DOCUMENT RESUME

ED 041 752 SE 009 204

AUTHOR	Bowers, Wayne A.
TITLE	Distributions.
INSTITUTION	Commission on Coll. Physics, College Park, Md.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	65
NOTE	48p.; Monograph written for the Conference on the New Instructional Materials in Physics (University of Washington, Seattle, 1965)
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.50
DESCRIPTORS	*College Science, *Instructional Materials, Measurement, *Physics, *Probability Theory, Resource Materials, *Statistics

ABSTRACT

This monograph was written for the Conference of the New Instructional Materials in Physics, held at the University of Washington in summer, 1965. It is intended for students who have had an introductory college physics course. It seeks to provide an introduction to the idea of distributions in general, and to some aspects of the subject in physics. There are three chapters. Chapter 1 gives a non-mathematical treatment of distributions. Chapter 2 considers means, other averages standard deviation, and the binomial distribution. Chapter 3 concerns continuous distributions and requires students to be familiar with elementary calculus. Problems are presented at the end of each chapter. (LC)

ED041752

FILE COPY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Distributions

COMMISSION ON COLLEGE PHYSICS
DEPARTMENT OF PHYSICS AND ASTRONOMY
UNIVERSITY OF MARYLAND
4321 HARTWICK ROAD
COLLEGE PARK, MD. 20740

WAYNE A. BOWERS

University of North Carolina

SE 009 204

GENERAL PREFACE

This monograph was written for the Conference on the New Instructional Materials in Physics, held at the University of Washington in the summer of 1965. The general purpose of the conference was to create effective ways of presenting physics to college students who are not preparing to become professional physicists. Such an audience might include prospective secondary school physics teachers, prospective practitioners of other sciences, and those who wish to learn physics as one component of a liberal education.

At the Conference some 40 physicists and 12 filmmakers and designers worked for periods ranging from four to nine weeks. The central task, certainly the one in which most physicists participated, was the writing of monographs.

Although there was no consensus on a single approach, many writers felt that their presentations ought to put more than the customary emphasis on physical insight and synthesis. Moreover, the treatment was to be "multi-level" --- that is, each monograph would consist of several sections arranged in increasing order of sophistication. Such papers, it was hoped, could be readily introduced into existing courses or provide the basis for new kinds of courses.

Monographs were written in four content areas: Forces and Fields, Quantum Mechanics, Thermal and Statistical Physics, and the Structure and Properties of Matter. Topic selections and general outlines were only loosely coordinated within each area in order to leave authors free to invent new approaches. In point of fact, however, a number of monographs do relate to others in complementary ways, a result of their authors' close, informal interaction.

Because of stringent time limitations, few of the monographs have been completed, and none has been extensively rewritten. Indeed, most writers feel that they are barely more than clean first drafts. Yet, because of the highly experimental nature of the undertaking, it is essential that these manuscripts be made available for careful review

by other physicists and for trial use with students. Much effort, therefore, has gone into publishing them in a readable format intended to facilitate serious consideration.

So many people have contributed to the project that complete acknowledgement is not possible. The National Science Foundation supported the Conference. The staff of the Commission on College Physics, led by E. Leonard Jossem, and that of the University of Washington physics department, led by Ronald Geballe and Ernest M. Henley, carried the heavy burden of organization. Walter C. Michels, Lyman G. Parratt, and George M. Volkoff read and criticized manuscripts at a critical stage in the writing. Judith Bregman, Edward Gerjuoy, Ernest M. Henley, and Lawrence Wilets read manuscripts editorially. Martha Ellis and Margery Lang did the technical editing; Ann Widditsch supervised the initial typing and assembled the final drafts. James Grunbaum designed the format and, assisted in Seattle by Roselyn Pape, directed the art preparation. Richard A. Mould has helped in all phases of readying manuscripts for the printer. Finally, and crucially, Jay F. Wilson, of the D. Van Nostrand Company, served as Managing Editor. For the hard work and steadfast support of all these persons and many others, I am deeply grateful.

Edward D. Lambe
Chairman, Panel on the
New Instructional Materials
Commission on College Physics

D I S T R I B U T I O N S

PREFACE

This monograph is intended to provide an introduction to the idea of distributions in general, and to some aspects of the subject of importance in physics. The level is intended to be suitable for students who have had an introductory college physics course, although only very little knowledge of physics is actually required. The first chapter is entirely nonmathematical, the second is intended to be understandable to students who have not had a calculus course, and the third chapter requires familiarity with elementary calculus. A fourth chapter is planned which treats applications in statistical mechanics, but that chapter is not included in the present publication.

It is my hope that the material may be helpful in giving a somewhat more detailed introduction to certain statistical and probabilistic ideas than commonly occurs in the standard physics texts, and that it may therefore find some use in preparing students not only for the study of kinetic theory and statistical mechanics, but also for other areas of physics where these ideas are useful.

Wayne A. Bowers

CONTENTS

1	INTRODUCTION	1
1.1	The Idea of a Distribution	1
1.2	Graphical Representation of Distributions	2
1.3	Cumulative Distributions	3
1.4	Joint Distributions	4
1.5	Probability Distributions; Fluctuations	5
1.6	Experiment and Theory	6
2	DISCRETE DISTRIBUTIONS	8
2.1	Notation; Mean and Other Averages	8
2.2	Width of a Distribution; Standard Deviation	9
2.3	The Binomial Distribution	10
2.4	The Asymmetric Binomial Distribution	15
2.5	The Multinomial Distribution	16
3	CONTINUOUS DISTRIBUTIONS	21
3.1	Introduction; Mean Values; Examples	21
3.2	Change of Variable	24
3.3	Radioactive Decay; Molecular Free Paths	28
3.4	Other Molecular Path Problems; Poisson Distribution	31
3.5	Gaussian Distribution; Errors	34
3.6	Joint Distributions; Maxwellian Velocity Distribution	36

1 INTRODUCTION

1.1 THE IDEA OF A DISTRIBUTION

Let us imagine ourselves capable of seeing individual molecules flying about in a gas. Suppose we fix our attention on a small region, and consider the motion of the molecules as they move across the region. Suppose we are asked: "How fast do they travel?" On looking at several molecules, we find that some travel slowly, some fast; perhaps we find a tendency for a certain range of speeds to predominate, but nevertheless values outside such a range occur from time to time.

Again we look at the gas; this time we follow an individual molecule in its path. It collides with another molecule, changes its direction and speed, goes on to another collision, strikes the wall of the vessel and bounces off, collides with still a third molecule - and so on. Suppose we are asked: "How far does it travel between collisions?" Again there is no single answer. Sometimes it travels a micron, sometimes much less; the distances vary widely.

Or again, suppose we are simply asked: "How many molecules are there in a cubic micron?" If there is a reasonably high vacuum in the system, there may be a density of about ten molecules per cubic micron; but again, as we watch any particular cubic micron of volume, we see now eight, now thirteen, now eleven, now seven molecules. The number we see varies "randomly" about an average value.

The questions asked about the molecules in each of the examples given have this one thing in common with each other and in common with a host of other questions which arise in physics and in the other natural sciences and the social sciences. They cannot be answered with a single number, but only with a whole range of numbers. In this respect they are in contrast with such questions as: "What is the speed of light in vacuo?" and "What is the temperature of pure boiling water under normal atmospheric pressure?", which have precise numerical answers (ignoring very small uncertainties which decrease further with each improvement in the experimental apparatus). The first type of question can be given only such answers as: "Out of 200 molecules, 52 had speeds between 0 and 300 meters/sec, 89 had speeds between 300 and 600 meters/sec, and the remaining 59 had speeds greater than 600 meters/sec."

This kind of an answer we call a distribution because it tells how the molecules are distributed with respect to the property of interest - speed, in this case. Such distributions arise in every area of physics. The time of decay of a radioactive nucleus, the angle of scattering of a neutron colliding with a carbon nucleus, the position of an electron in a hydrogen atom, the energy of the beta particle emitted in the radioactive decay of a nuclear species - all of these and a host of others are described by distributions rather than by single numbers. In the social sciences, perhaps even more than in physics, distributions are ubiquitous; such familiar examples as the distributions of income, of life expectancy, or of education come to mind. Not only is the determination of such distributions the aim of much research in the social sciences; once determined, they form the essential factual base for further economic and sociological work.

To define a distribution we must first specify a population (molecules of oxygen gas at normal pressure and temperature; two-million-volt neutrons scattered from carbon nuclei; students in a certain physics course), and a characteristic or property of the individuals comprising the population

which can be measured (speed of the molecule; angle of scattering of the neutron; final examination grade of the student). A table, a graph, or a mathematical function telling how many of the population have specified values of the property in question then constitutes the distribution. More explicitly, it is sometimes called a frequency distribution, since it gives the frequency of occurrence of the specified values of the property in question.

We will distinguish between discrete and continuous distributions. By a discrete distribution we will mean one for which a finite number of categories are used for specifying the property in question. This may happen in two ways. First, the property may be intrinsically discrete, as in the example described earlier of the number of molecules in a cubic micron, which is necessarily an integer. (Many distributions arising in probability theory are of this type. The number of heads in a sequence of coin tosses and the number arising in the throw of a pair of dice have such distributions.) Second, the property in question, although having in principle a continuous range of values, may be divided into a finite number of intervals for convenience. Thus the range of scattering angles for the neutron extends continuously from 0° to 180°, but it may be divided into eighteen 10° intervals or thirty-six 5° intervals for specifying the observed distribution. The scale of such a division may be determined in part by the instrumental limitations (perhaps intervals less than 5° cannot be accurately defined by the particular apparatus), and in part by the amount of data accumulated.

By a continuous distribution, we will mean one for which the full continuous range of values of the property is used, in the sense that the intervals which would characterize a discrete distribution are allowed to become arbitrarily small. To describe adequately how this is done, the methods of the calculus must be used. We will therefore postpone detailed discussion of continuous distributions to Chapter 3. We may remark, however, that in principle an infinite population would be needed to specify a continuous distribution, for as the intervals are taken smaller, their number increases. Thus with any finite population, there is a limit to the possible decrease in interval size. Nevertheless, the populations which enter in many physical problems are so enormous (as for example the 10^{19} molecules in a cubic centimeter of gas) that they are effectively infinite, and the methods of continuous distributions may be used without trouble.

1.2 GRAPHICAL REPRESENTATION OF DISTRIBUTION

Discrete distributions may be represented graphically in various ways. We shall use two slightly different methods, one of which is appropriate for the "intrinsically discrete" distributions discussed in the previous

NUMBER OF HEADS	FREQUENCY OF OCCURRENCE
0	2
1	14
2	19
3	11
4	4
	50

Table 1.1

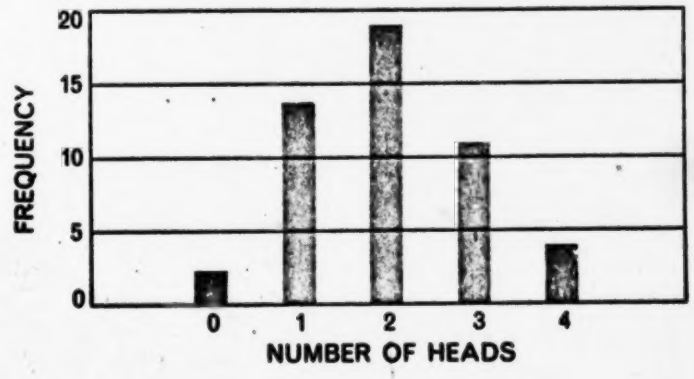


Fig. 1.1

VELOCITY INTERVAL, METERS PER SECOND	NUMBER OF MOLECULES WITH GIVEN VELOCITY
0 — 100	4
100 — 200	16
200 — 300	35
300 — 400	44
400 — 500	37
500 — 600	28
600 — 700	17
700 — 800	11
800 — 900	6
900 —1000	2
	200

Table 1.2

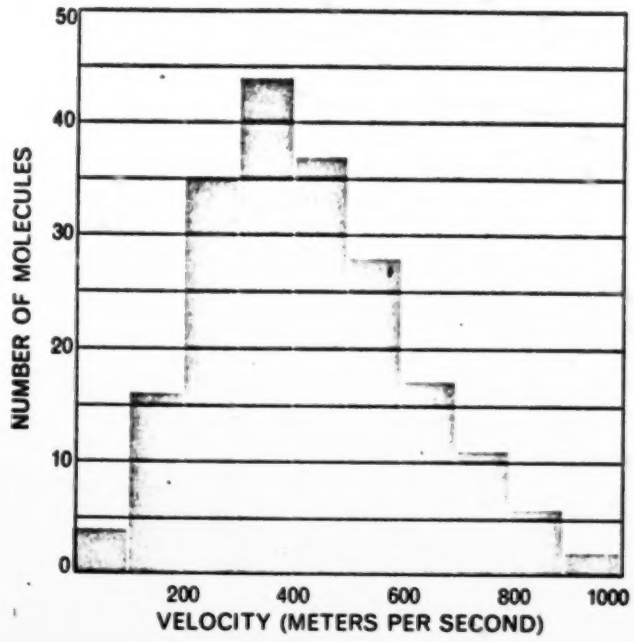


Fig. 1.2

section, and the other for the distributions of continuously variable quantities whose range is divided into intervals. For the former we use a bar graph, for the latter a histogram. Each is essentially a plot of frequency of occurrence vertically against the possible values of the quantity horizontally. The bar graph uses vertical lines at the positions of the discrete index (usually an integer), since values between the discrete indices are meaningless. The histogram uses rectangles of the appropriate heights erected on each interval, since the individual values may have occurred anywhere in the interval into which they have been grouped.

We will give an example of each type of graphical representation. In Table 1.1 and Fig. 1.1, we have a table, and the corresponding bar graph, giving the frequency of occurrence of various numbers of "heads" in a series of fifty tosses of a group of four coins. The possible outcomes are of course 0, 1, 2, 3, or 4 heads in each toss, so a bar graph is appropriate. In Table 1.2 and Fig. 1.2, we have a table and the corresponding histogram giving the distribution of speeds, in intervals of 100 meters/sec, of 200 molecules.

Although it contains no more information than the table from which it is constructed, the bar graph or histogram is useful in giving a quick qualitative impression of a distribution. It can readily give a visual comparison of two distributions, or of an observed distribution with a theoretical or calculated one. But for more quantitative information, such as averages and other numbers associated with the distribution, one must usually refer to the data in the table.

1.3 CUMULATIVE DISTRIBUTIONS

Occasionally a slightly different arrangement of the same information is useful. We can, instead of giving the numbers of molecules in each velocity interval as in Table 1.2, give the total number of molecules with velocity less than 100, 200, ... meters/sec. Such a specification is known as a cumulative distribution. In Table 1.3 and Fig. 1.3 (see next page), the same data as in Table 1.2 are given in this fashion. Evidently the two types of distribution are easily obtained from one another; for example, the differences between successive entries in the table for the cumulative distribution give the entries for the corresponding intervals of the original fre-

VELOCITY METERS PER SECOND	NUMBER OF MOLECULES WITH LESS VELOCITY
0	0
100	4
200	20
300	55
400	99
500	136
600	164
700	181
800	192
900	198
1000	200

Table 1.3

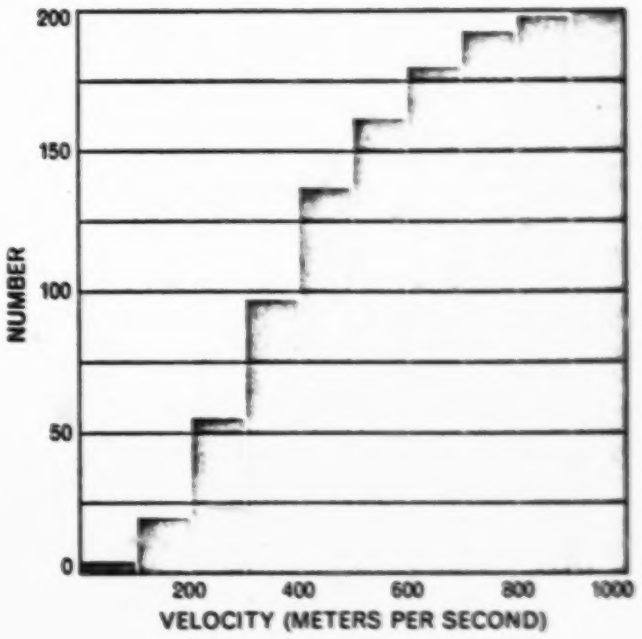


Fig. 1.3

quency distribution. Notice that a cumulative distribution cannot decrease as one goes up the scale of the measured property. In fact, it must ultimately increase from zero to the number in the population.

A different way of specifying a cumulative distribution is familiar in the treatment of test scores. If a student is told he stands in the "78th percentile" on a certain test, he knows that his score is higher than that of 78% of the students taking the test. This language (percentiles, or deciles, or quartiles) corresponds to using equal intervals (hundredths, tenths, quarters) of the total range of the vertical, or number, axis of such a graph as that of Fig. 1.3, rather than of the horizontal axis. One is asking, in effect, not how many molecules have velocities lying in the various equal intervals, but rather what intervals of velocity correspond to equal numbers of molecules - one percent, or one tenth, or one quarter, respectively, of the total number.

1.4 JOINT DISTRIBUTIONS

In all the examples cited so far, one property of the individuals comprising the population has been singled out for attention: the speed of the molecules, the grades of the students, the angle of scattering of the neutrons. But the individuals also have other properties. The molecules have position and direction of motion as well as speed; the students have ages, heights, and blood pressures as well as exam grades; the neutrons have energies and momenta as well as angles of scattering. The various properties may be related to one another, or they may be quite independent. The molecule's speed - but not its direction! - is closely related to its kinetic energy; the student's height is unrelated to his examination grade, but is somewhat related to his weight. In either case, we may define a joint distribution of two or more such characteristics. By this we mean a listing, by table, or graph, or mathematical function, of the number of individuals of the population that have simultaneously certain specified values of each of the two or more characteristics.

We give in Table 1.4 and Fig. 1.4 a rather prosaic example: a joint distribution of heights and weights of 5000 men. Notice that a two-dimensional array rather than a column is needed for the table, and a three-dimensional histogram for the figure. The difficulties of pictorially representing joint

distributions for more than two characteristics are apparent! Nevertheless they are of great importance in physics. For example, the joint distribution in position and speed of a molecule, and the joint distribution of the three components of the velocity of a molecule, are basic distributions in the kinetic theory of gases.

1.5 PROBABILITY DISTRIBUTIONS; FLUCTUATIONS

Each of the frequency distributions we have studied may be converted to a relative frequency distribution by dividing each entry in the table specifying the distribution by the total number in the population. We will thus obtain a table consisting of fractions whose sum is unity. Each entry in the new table will give the fraction of the total population which lies in the specified interval. Now suppose we make another set of measurements of the same kind on a new population of the same type, whose total number is not the same. The resulting frequency distribution will of course be different from the first; but if we again convert to relative frequency, we expect the fractions in the new distribution to be not too different from those in the original set. Experience shows that as we accumulate more and more data corresponding to larger and larger populations, the fractions giving the relative frequencies of the various alternatives tend to approach limits. These limiting values we call the probabili-

HEIGHT \ WEIGHT	125–150 LBS.	150–175 LBS.	175–200 LBS.	200–225 LBS.	225–250 LBS.	TOTALS
5′ –5′3″	9	22	19	1	0	51
5′3″–5′6″	46	158	168	76	27	475
5′6″–5′9″	51	227	423	332	89	1122
5′9″–6′0″	62	309	768	683	197	2019
6′0″–6′3″	21	212	392	453	120	1198
6′3″–6′6″	0	9	37	72	17	135
TOTALS	189	937	1807	1617	450	5000

Table 1.4

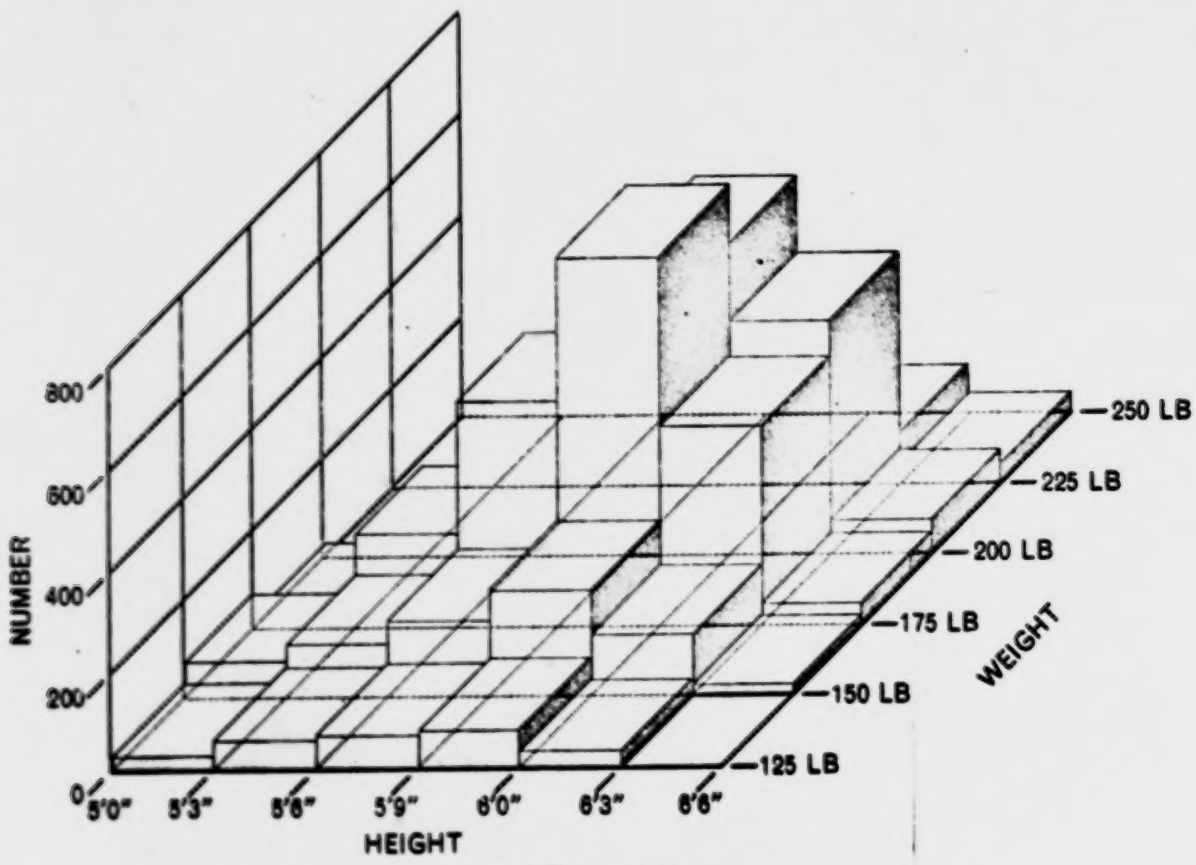


Fig. 1.4

ties of the various alternatives, and the whole set of probabilities for all the alternatives, which must add up to unity, we call a probability distribution. Evidently such a definition can only make sense for those cases in which populations of sufficiently large numbers actually exist. The examples from physics with which we are chiefly concerned are of this kind; the same is not always necessarily so in other contexts. It would be difficult, for example, to attach much meaning to the statement: "The probability that billion-dollar corporations will go bankrupt in 1966 is 0.02%"; but a similar assertion about small companies, of which there are many thousands, could be perfectly sensible.

So long as we are dealing with a "sufficiently large" sample of the over-all population of interest, we may identify its relative frequency distribution with the probability distribution of the population. The work of the next chapter will lead us to a criterion for judging whether a sample is "sufficiently large." But what if the sample is not that large? Clearly, different samples will exhibit somewhat different distributions, even though they are drawn from the same larger population. These variations (in physics they are often called "fluctuations") are of great importance in practical applications. An opinion-polling organization must know how large a sample to question in order that the results obtained are reasonably representative of the actual distribution of opinion in the whole population. A manufacturer producing large numbers of standardized items needs to know how large a sample must be tested for conformity to the standard, in order to have reasonable assurance that no defective units are allowed on the market. Here again, the work of the next chapter will give us means of estimating these variations for samples of given sizes.

1.6 EXPERIMENT AND THEORY

In most of the examples discussed above, we have been thinking of the distributions as being found experimentally. A large number of molecular velocities, or neutron-scattering angles, or men's heights and weights, are measured, tabulated, and converted to distributions. But not all distributions are experimental; they may be deduced theoretically from physical or mathematical assumptions. Frequently they require techniques from various branches of mathematics - particularly probability theory - for their derivation. Since much of probability theory is concerned with the calculation of other probabilities from sets of given ones, the basic rules for combining probabilities are used frequently in the following chapters. We will therefore state them here for reference.

First: if A and B are mutually exclusive alternatives, the probability that either A or B will occur is the sum of the probabilities of A and of B occurring separately. (Example: the probability that either 1 or 2 heads show in a toss of 4 coins is the sum of the probabilities that one head shows, and that two heads show.)

Second: The probability that first A and then B occur in successive independent trials is the product of the probabilities of A and of B occurring separately. (Example: the probability that first two heads show, and then one head shows in a second toss of four coins, is the product of the probabilities that one head shows and that two show.)

With the help of these apparently simple rules, elaborate superstructures of theory can be erected. But it is important to remember that, for the physicist, the ultimate test of the validity of theoretical calculations of the type which we will do in the next chapters is their comparison with experiment.

PROBLEMS

1.1 Give examples of a distribution which might be of interest in (a) psychology, (b) linguistics, (c) economics. In each case, specify the population, and the range of values of the characteristic whose distribution you envisage.

1.2 Sometimes a cumulative distribution is defined by the number greater than (instead of less than) a series of successive equally spaced values of the property in question. Construct such a cumulative distribution for the molecular velocities of Table 1.2.

1.3 What is the meaning of the last row and of the last column (labeled totals) of the joint distribution given in Table 1.4?

1.4 The median of a distribution is the value of the property such that half the population lies above and half below it in value. What is the median for each of the distributions in Table 1.1 and Table 1.2?

1.5 Taking the distribution of velocities in Table 1.2 to be sufficiently representative of the whole population, what is the probability that: (a) the velocity of a molecule lies between 200 and 700 meter/sec? (b) that three molecules chosen independently all have velocities lying between 300 and 400 meters/sec?

1.6 Imagine an instructor who gives only three marks, high (H), medium (M), and low (L), on every test, and who gives one third of the class each mark. When averaging two tests, he gives H only to those who have H on each, and similarly for L. All the rest get M. What fraction of the class will have averages of H, L, and M, respectively, on two tests? On three? Using the foregoing instance as a guide, explain why a student who is never first in the class on any given test, but who is consistently high in standing, often ends the term with the best average in the class.

2 DISCRETE DISTRIBUTIONS

2.1 NOTATION; MEAN AND OTHER AVERAGES

We will need a notation to describe the discrete distributions with which this chapter will deal. Let us denote by N the total number in the population in question; then the number of individuals falling in the kth interval of the property whose distribution is under study we will denote by n_k. The sum of all the numbers in the various groups must be N; using the customary notation for summation, we have

$$\sum_k n_k = N, \qquad (2.1)$$

where the summation extends over all the intervals into which the range of values has been divided. The corresponding probabilities we will denote by p_k; as remarked in Section 1.6, this probability is found by dividing n_k by N:

$$p_k = n_k/N, \qquad (2.2)$$

or

$$n_k = N\, p_k,$$

from which we obtain

$$\sum_k p_k = \frac{1}{N} \sum_k n_k = \frac{1}{N} \cdot N = 1, \qquad (2,3)$$

Thus the probability that an individual chosen at random from the population is found to be in the kth interval is p_k; Eq. (2.3) represents the statement that the total probability of finding the individual somewhere among all the intervals is unity, i.e., it is sure to be found in one or another interval.

The value of the property under study, in the kth interval, will be denoted by an appropriate symbol - differing from example to example - with subscript "k"; in the example of heights in Table 1.4, we may use h_k, for velocities of molecules, in Table 1.2 we might use v_k, and so on. Notice that there is some ambiguity in the phrase "the value in the kth interval"; there is no unique "value" of the height for the interval from 5 ft 6 in. to 5 ft 9 in. In examples of this kind we will agree to use the value at the midpoint of the interval (e.g., 5 ft $7\frac{1}{2}$ in. for the case cited). In certain other examples of discrete distributions there will be no ambiguity, however, because the characteristic studied takes on strictly discrete (perhaps integer) values. If we ask for the distribution of the total obtained in a number of throws of a pair of dice, for instance, the values are integers ranging from 2 to 12; in examples of this type, the phrase "value in the kth interval" may be replaced throughout by "kth value" in the preceding discussion.

An extremely important quantity associated with the distribution is the mean value or average of the property under study. It is defined just as the ordinary arithmetic average of common usage, as the sum of the values of the property for all individuals of the population, divided by the total number in the population. Since the individuals are grouped in such a way that n_1 have the value h_1, n_2 have the value h_2, and so on, the average, which we will denote by $\langle h \rangle$, is given by

$$\langle h \rangle = (n_1 h_1 + n_2 h_2 + \ldots)/N$$

$$= (\sum_k n_k h_k)/N. \qquad (2.4)$$

Bearing in mind Eq. (2.2), which defined the probabilities p_k, we may also write

$$\langle h \rangle = \sum_k p_k\, h_k. \qquad (2.5)$$

This mean value is the single number most often used to characterize a distribution - if a single number must

be used! In everyday usage this is very familiar; one hears references to such things as the average income in a state or country, the average life expectancy of a 25-year-old female, or the average number of years of schooling of this group or that. In each of these cases, the "average" is being used as a single number to characterize a whole body of information which really constitutes a distribution. The actual incomes may vary from 0 to $100,000, the life expectancies from 1 to 75 years, and the years of education from 0 to 20; but the average - which is the single value which all the individuals would need to have, in order to give the same total value for the whole population that in fact exists - is used as a quick summary.

It is however by no means the only mean or average which may be needed. There are many other averages which may be defined, and which may be more appropriate for some particular use. The molecules of a gas have a distribution of speeds (the "Maxwellian" distribution, which we will study later); and the average or mean speed defined as in Eqs. (2.4) or (2.5):

$$\langle v \rangle = \sum_k p_k v_k = (\sum_k n_k v_k)/N$$

is often useful. But the kinetic energy of a molecule, according to classical mechanics, is proportional to the square of the velocity; hence if we need (as we will in the kinetic theory of gases) information about the average, or mean, kinetic energy of the gas of molecules, we must find the average of the square of the velocity:

$$\langle v^2 \rangle = \sum_k p_k v_k^2 = (\sum_k n_k v_k^2)/N. \qquad (2.6)$$

In a similar fashion, other functions of the velocity which occur in other contexts may need to be averaged over the distribution of velocities, and we will use a similar notation in each case:

$$\langle f(v) \rangle = \sum_k p_k f(v_k) = (\sum_k n_k f(v_k))/N. \qquad (2.7)$$

In particular, the averages of the powers of the quantity whose distribution is being studied, which are known as the moments of the distribution, are of interest; not infrequently in physics, various moments of a distribution may be accessible to direct measurement, even though the distribution as a whole is not. From a knowledge of the moments one can reconstruct - approximately - the distribution itself, or at least confirm whether or not some theoretically predicted distribution yields the observed moments.

2.2 WIDTH OF A DISTRIBUTION; STANDARD DEVIATION

If the mean value is the single number most often used to characterize a distribution, the number second in importance after the mean is one which measures the width of the distribution, or the spread of values about the mean. Clearly, many widely different distributions can have the same mean; Figs. 2.1, 2.2, and 2.3 give examples of pos-

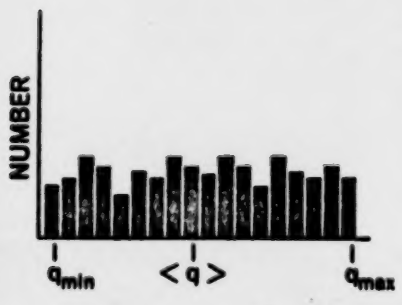


Fig. 2.1

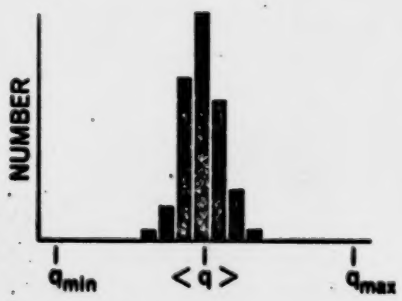


Fig. 2.2

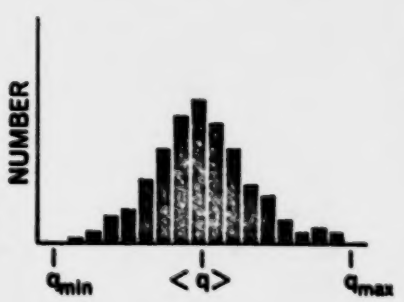


Fig. 2.3

sible distributions of a quantity q whose values range from q_{min} to q_{max}. In each case the mean value $\langle q \rangle$ is the same, but the "shapes" of the distributions are quite different; in Fig. 2.1, all values between q_{min} and q_{max} are nearly equally likely, in Fig. 2.2, only values near the mean are likely, and Fig. 2.3 is intermediate between the extremes represented by the first two.

How is the "width" to be measured? Evidently what is needed is some estimate of the likelihood of various deviations from the mean value; large deviations are likely in Fig. 2.1, unlikely in Fig. 2.2, and moderately likely in Fig. 2.3. One's first thought might reasonably be to take the average of the deviations from the mean. Denoting by Δq_k the difference between a particular value q_k and the mean $\langle q \rangle$

$$\Delta q_k = (q_k - \langle q \rangle). \qquad (2.8)$$

we can calculate its mean in the standard way:

$$\langle \Delta q \rangle = \langle (q - \langle q \rangle) \rangle = \left[\sum_k n_k (q_k - \langle q \rangle) \right] / N$$
$$= \sum_k p_k \, (q_k - \langle q \rangle); \qquad (2.9)$$

but a little reflection shows that this quantity is necessarily zero. It is customary therefore to take instead the mean of the squared deviation from the mean:

$$\langle (\Delta q)^2 \rangle = \langle (q - \langle q \rangle)^2 \rangle$$
$$= \left[\sum_k n_k \, (q_k - \langle q \rangle)^2 \right] / N. \qquad (2.10)$$

This quantity is called the variance of the distribution, and its square root (that is, the "root mean square" deviation) is called the standard deviation of the distribution. It is often denoted by σ (lower case Greek "sigma"):

$$\sigma = \sqrt{\langle (q - \langle q \rangle)^2 \rangle} = \sqrt{\sum_k p_k \, (q_k - \langle q \rangle)^2}$$
$$= \sqrt{\sum_k p_k \, (\Delta q_k)^2}. \qquad (2.11)$$

A useful alternative expression for σ may be found by expanding the square deviation:

$$\sigma^2 = \sum_k p_k \, (\Delta q_k)^2$$
$$= \sum_k p_k \, (q_k^2 - 2\langle q \rangle q_k + \langle q \rangle^2)$$
$$= \langle q^2 \rangle - 2\langle q \rangle \langle q \rangle + \langle q \rangle^2$$
$$= \langle q^2 \rangle - \langle q \rangle^2. \qquad (2.12)$$

The last steps have made use of Eqs. (2.3), (2.5), and (2.6). In words, one may say that the variance is the difference between the mean square of the quantity and the square of the mean of the quantity. Since, in the language used in the last part of the previous section, the mean is the "first moment" and the mean square is the "second moment" of the distribution, one may say that knowledge of value of the first moment specifies the mean, and of the second moment the variance. Higher moments would give successively more detailed features of the distribution. For example, the third moment will give some indication of whether positive or negative deviations from the mean predominate; the variance gives no clue to this, since it involves the square of the deviation, to which positive and negative deviations contribute equally.

2.3 THE BINOMIAL DISTRIBUTION

A distribution which arises naturally in a variety of physical problems is the binomial distribution. It arises whenever a choice of two alternatives is available, and the choice is made many times. For example, in a "random walk" problem, which is a model for Brownian motion or for molecular diffusion, one imagines a particle moving to the right or left along a line in steps of equal size. If each step is equally likely to be to the left or to the right, what is the likelihood that the particle will have moved a certain net distance to the right after a given number of steps? Or again, under certain conditions, an atom with a magnetic moment in a magnetic field may,

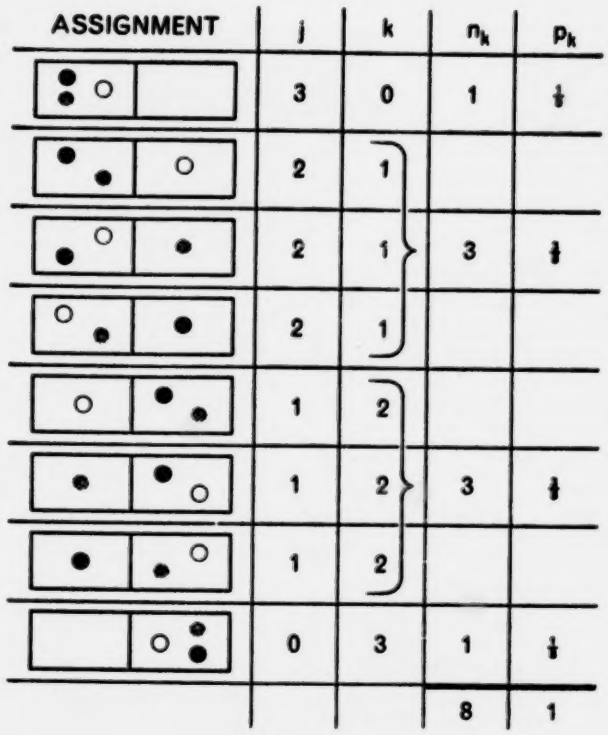
according to quantum theory, align itself in two ways only: along or against the field. In a large collection of such atoms, what is the likelihood of finding a given number aligned with or against the field? An even simpler question which we will adopt for illustrating the binomial distribution, is the following: In a box of gas, how many of the molecules will be at any instant in the right half of the box? The question may sound trivial at first; surely half the molecules are in each half of the box - at least "on the average"! But how do we know that this is so? Or even if it is so, what does "on the average" mean? What are the chances of finding a few more or a few less than half of the molecules in the right half? Does it matter whether the gas is at normal pressure or extremely rarefied?

To consider this problem, suppose the box contains M molecules; let j be in the left half and k in the right half. Then we must have

$$j + k = M. \qquad (2.13)$$

The possible values of k range from 0 to M; we want the distribution of k over these values. To illustrate the method, consider first the case of three molecules, although our interest is really in large numbers. In this case we can simply enumerate the possible ways of assigning molecules to the two halves respectively. In Fig. 2.4, the possible assignments are sketched, and the numbers j and k for each assignment are listed, together with the number n_k of assignments for which k molecules are in the right half, and the probability p_k of finding k molecules in the right half.

Figure 2.4 shows that there are a total of eight possible assignments. This is understandable, since there are two possibilities for each molecule, hence $2 \times 2 \times 2$ for all three. Of these eight assignments, one corresponds to no molecules in the right half, three correspond to one molecule, three to two molecules, and one to three. The



ASSIGNMENT	j	k	n_k	p_k
	3	0	1	1/8
	2	1		
	2	1	3	3/8
	2	1		
	1	2		
	1	2	3	3/8
	1	2		
	0	3	1	1/8
			8	1

Fig. 2.4

corresponding probabilities are respectively 1/8, 3/8, 3/8, and 1/8, as indicated in the last column of the figure. Thus, in a very large number of boxes containing three molecules, we expect to find that 1/8 of them contain no molecules in the right half, 3/8 of them contain one, 3/8 of them contain two, and 1/8 of them contain three. Alternatively, we may say that these fractions give the number of times, in a sequence of a large number of independent looks at the same box, that one will see the specified number of molecules in the right half. From these probabilities we may calculate the mean number, and the various other averages which might be of interest, in accordance with the formulas of sections 2.1 and 2.2. Instead of doing this for the special case of three molecules, however, let us first go on to the general case of M molecules.

In the case of M molecules, there will be a total of 2^M possible assignments of each of the M molecules to the right or the left half. How many of these correspond to precisely k molecules in the right half? Only one assignment gives k = 0: every molecule on the left. But M assignments give k = 1; the single molecule on the right

may be chosen in M different ways. For k = 2, the first of the two molecules on the right may be chosen in M ways, and the second in (M − 1) ways; this gives a total of M(M − 1) ways. But notice that this counts as distinct assignments those in which the same two molecules are placed on the right, but in reversed order. Since a single assignment is specified by saying, for example, "put molecules #5 and #17 on the right" without taking into account the order in which #5 and #17 are selected, we must correct for the spurious doubling of the number of assignments which our method of counting gives. Taking this into account gives finally the result M(M − 1)/2 for the number of distinct assignments of two molecules to the right half. For k = 3, the same line of reasoning gives the result M(M − 1)(M − 2)/6; here the factor in the numerator gives the number of ways of picking three molecules in a particular order, and the factor 6 in the denominator corrects for the number of permutations of the three molecules selected to be in the right half. For general k, the result is

$$n_k = M(M-1)(M-2)\cdots(M-k+1)/k!, \qquad (2.14)$$

where again the denominator contains the factor k! (to be read "k factorial," the product of the first k integers), which is the number of permutations of k objects, and which corrects for the overcounting of the assignments which the numerator effects. A somewhat more symmetrical form can be found by multiplying numerator and denominator both by (M − k)!, the product of the integers from 1 up through (M − k):

$$n_k = \frac{M!}{k!\,(M-k)!}. \qquad (2.15)$$

Bearing in mind Eq. (2.13), we may also write:

$$n_k = \frac{M!}{k!\,j!}, \quad (k + j = M). \qquad (2.16)$$

In this form, the symmetry between right and left halves becomes apparent; interchanging k and j leaves the expression unchanged, as it should. One can now easily verify that the expressions we wrote down earlier for n_1, n_2, and n_3 agree with the general expression. For n_0,, which we saw earlier has the value 1, the general expressions in Eqs. (2.15) or (2.16) are valid, provided we adopt the customary convention that 0! = 1.

The name "Binomial Distribution" for our result arises from the fact that the numbers n_k defined by Eq. (2.16) are precisely the same numbers that occur as the coefficients in the binomial expansion of mathematics:

$$(a + b)^M = \sum_{k=0}^{M} n_k\, a^{M-k} b^k. \qquad (2.17)$$

That this is so is understandable when one reflects that the kth coefficient in the expansion counts the number of ways of picking out k b's and (M − k) a's from the set of M factors (a + b) that are implied by $(a + b)^M$. Formally, this is identical with our problem of picking out k molecules to put in the right half and (M − k) to put in the left half, from the total of M molecules. We can use Eq. (2.17) to verify that the total number of assignments is 2^M, as we asserted earlier; for if we

M \ k	0	1	2	3	4	5	6	7	8	$\sum_{k=0}^{M} n_k$
1	1	1								2
2	1	2	1							4
3	1	3	3	1						8
4	1	4	6	4	1					16
5	1	5	10	10	5	1				32
6	1	6	15	20	15	6	1			64
7	1	7	21	35	35	21	7	1		128
8	1	8	28	56	70	56	28	8	1	256

Table 2.1

let a = b = 1 in Eq. (2.17), we get

$$(1 + 1)^M = \sum_{k=0}^{M} n_k (1)^{M-k} (1)^k,$$

or

$$2^M = \sum_{k=0}^{M} n_k = \sum_{k=0}^{M} \frac{M!}{k!\ (M-k)!} \quad (2.18)$$

In Table 2.1, the binomial coefficients are listed for values of M from 1 to 8; for each M, the values of k run from 0 through M. The sum of the coefficients is also listed for each M. Notice that the table can be very easily continued; given the set of n_k for a particular M, the next row in the table, corresponding to M + 1, is given by the rule:

$$n_k\ (M+1\ \text{row}) = n_k + n_{k-1}\ (M\ \text{row}) \quad (2.19)$$

The corresponding probabilities are found by dividing the n_k by the total number of assignments 2^M; hence

$$p_k = \frac{M!}{2^M\ k!\ (M-k)!}. \quad (2.20)$$

We can now examine the distribution and use it to answer some of the questions raised initially. A glance at Table 2.1 shows that, for all values of M occurring there, the most probable value of k (the one for which n_k, and hence also p_k is largest) is always M/2, if M is even, and that if M is odd, the two integers nearest to M/2 are equally the most probable. It is not hard to show that this remains true for any M. Thus one is indeed more likely to find just one half of the molecules in the right (or the left) half than any other particular value. What about the average, or mean number? From the definition, it is given by

$$\langle k \rangle = \sum_{k=0}^{M} k p_k = \frac{1}{2^M} \sum_{k=0}^{M} \frac{kM!}{k!\ (M-k)!}. \quad (2.21)$$

By actual calculation, using the first few values of M and Table 2.1, one finds the value M/2 in each instance; that it is true in general follows from Eq. (2.21) by rewriting it slightly differently and using Eq. (2.18), with (M − 1) substituted for M:

$$\langle k \rangle = \frac{1}{2^M} \sum_{k=1}^{M} \frac{M(M-1)!}{(k-1)!\left[(M-1)-(k-1)\right]!}$$

$$= \frac{M}{2^M} \sum_{q=0}^{M-1} \frac{(M-1)!}{q!\ (M-1-q)!}$$

$$= \frac{M}{2^M} \cdot 2^{M-1} = M/2. \quad (2.22)$$

Hence the average number found in either half of the box, in many trials, will also be just half of the total number of molecules.

Finally, the likelihood of deviations from the mean value can be examined; for this purpose we need, as shown in section (2.2), the mean of the square of k. Before examining this quantity mathematically, let us see qualitatively what to expect, by plotting the distribution of k for various values of M. Fig. 2.5 (next page) shows bar graphs of n_k against k for M = 8, 40, 200, and 1,000. For ease of comparison they are drawn with the same ordinate at the maximum, and with the same range of the abscissa corresponding to the full scale from 0 to M in each case. For M = 200 and M = 1000, not every value of k has its n_k drawn in because of the smallness of the scale.

Notice that the graphs of the distributions become narrower as M increases. Since they are drawn to the same relative scale, this means that the probability of a deviation from the mean value M/2 by any given fraction of M becomes smaller as M increases. Thus although the probability of finding 3 or 5 molecules out of 8 in either half of the box is not very much less than the probability of finding 4, the probability of finding 15 or 25 out of

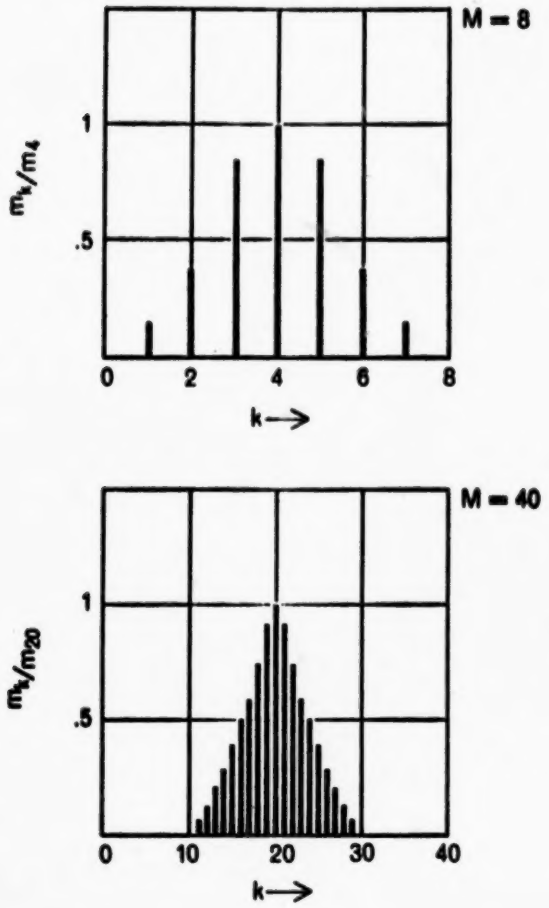


Fig. 2.5

40 is substantially less than that of finding 20. For M = 200, the corresponding numbers, 75 and 125, have extremely small probability compared with 100, and for M = 1000, the corresponding probabilities are entirely negligible.

We can confirm this qualitative conclusion by calculating the standard deviation, using the definition of section 2.2. According to Eq. (2.12), we must first find $\langle k^2 \rangle$:

$$\langle k^2 \rangle = \sum_{k=0}^{M} k^2 \, p_k$$

$$= \sum_{k=0}^{M} \frac{k^2 \, M!}{2^M \, k! \, (M-k)!}.$$

A trick which is useful, because of the occurrence of the factorial k in the denominator, is to rewrite k^2 as $(k^2 - k + k)$, or $(k\,(k-1) + k)$; then we have:

$$\langle k^2 \rangle = \langle k(k-1) + k \rangle = \langle k(k-1) \rangle + \langle k \rangle$$

$$= \sum_{k=2}^{M} \frac{k(k-1)\, M!}{2^M \, k! \, (M-k)!} + \langle k \rangle$$

$$= \frac{M(M-1)}{2^M} \sum_{k=2}^{M} \frac{(M-2)!}{(k-2)!\,[(M-2)-(k-2)]} + M/2$$

$$= \frac{M(M-1)}{2^M} \cdot 2^{(M-2)} + M/2$$

$$= (M^2 - M)/4 + M/2$$

$$= (M^2 + M)/4. \qquad (2.23)$$

Hence the standard deviation, according to Eq. (2.12), is qbtained from

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$$

$$= (M^2 + M)/4 - (M/2)^2$$

$$= M/4, \qquad (2.24)$$

and its ratio to the mean value $\langle k \rangle$, is:

$$\sigma/\langle k\rangle = (M/4)^{1/2}/(M/2) = 1/M^{1/2}\,. \qquad (2.25)$$

That is, the width of the distribution, although it increases in proportion to $M^{1/2}$ in absolute size, decreases in proportion to $1/M^{1/2}$ relative to the average number, in agreement with our qualitative result above based on the appearance of the graphs of Fig. 2.5. Thus a few cubic centimeters of gas, containing, say, 2×10^{20} molecules, will have "on the average" 10^{20} molecules in each half; this number may fluctuate by about 10^{10} from its mean value. Although 10^{10} is a very large number, it is a minute fraction of 10^{20}: one part in 10^{10}! It would be extremely hard ever to observe a deviation from the average so small as this. If however we imagine reducing the pressure until there are only 200 molecules in the same space, the average number of 100 in each half could fluctuate by about $\sqrt{100}$, or 10; this is 10%, or a sizable fraction of the average.

2.4 THE ASYMMETRIC BINOMIAL DISTRIBUTION

An assumption was hidden in the work of the previous section: plausible, but nevertheless an assumption. The eight alternatives listed explicitly in Fig. 2.4, for the assignment of three molecules to the two halves of the box, were regarded as "equally likely" to occur. But suppose the imaginary partition dividing the box into two parts is moved to the left, so that the left and right sections contain one-third and two-thirds of the total volume, respectively. The enumeration of assignments of Fig. 2.4 is still correct; but we feel it absurd to regard them all as "equally likely." How should we weight the various assignments?

It seems intuitively plausible to regard a single molecule as twice as likely to be found in the right side as the left under these circumstances, since its volume is twice as great. Stated otherwise, the probability of finding a single molecule on the left is one-third, and on the right is two-thirds. Then the probability for finding k molecules on the right and $(M - k)$ on the left is given by the number of assignments found previously, but multiplied by a factor $(2/3)^k \times (1/3)^{M-k}$, since probabilities of independent events are multiplied to find the probability of simultaneous occurrence. The result is therefore:

$$p_k = \frac{M!}{k!\,(M-k)!}\,(2/3)^k(1/3)^{M-k}. \qquad (2.26)$$

We can obtain the same result in a slightly different way. If a single molecule is twice as likely to be found on the right as on the left, a pair of molecules is four times as likely to be found on the right as on the left, and a triplet is eight times as likely. Hence the set of n_k's of Fig. 2.4 should be multiplied by 1 for $k = 0$, by 2 for $k = 1$, by 4 for $k = 2$, and by 8 for $k = 3$ to give the properly weighted assignments, as shown in Table 2.2 below. Notice that the total number of assignments with the new weights is 27, which is 3^3; dividing by this number, we obtain the set of probabilities p_k given in the last column. But these agree exactly with the result of Eq. (2.26), when M is set equal to 3.

More generally, if we divide the volume into two parts V_L and V_R, we may take the probability of finding a single molecule on the right to be

k	OLD n	WEIGHT	NEW n	NEW p
0	1	1	1	1/27
1	3	2	6	6/27
2	3	4	12	12/27
3	1	8	8	8/27
			27	1

Table 2.2

given bv $p = V_R/(V_L + V_R)$, and the probability of finding it on the left to be given by $q = V_L/(V_L + V_R)$. Then the same argument used in arriving at Eq. (2.26) leads to the following expression for finding k molecules out of a total of M to be in the volume V_R:

$$p_k = \frac{M!}{k!\ (M-k)!}\ p^k\ q^{M-k} \qquad (2.27)$$

From the binomial expansion, Eq. (2.17), we see that the sum of the p_k's is unity, as it should be:

$$\sum_{k=0}^{M} p_k = \sum_{k=0}^{M} \frac{M!}{k!\ (M-k)!}\ p^k\ q^{M-k}$$

$$= (p + q)^M = 1^M = 1.$$

The case discussed in section (2.3) is then the symmetric case $p = q = \frac{1}{2}$.

The general asymmetric binomial distribution defined by Eq. (2.27) applies whenever one makes M independent repetitions of a choice between two mutually exclusive alternatives whose probabilities are p and q, with $p + q = 1$. Other examples, in addition to the one used above, can easily be constructed from the kinds of physical problems discussed at the beginning of section 2.3. The biased random walk, in which the particle moves to the right with probability p, and to the left with probability q at each step, is such an instance.

Using the same methods as in the previous section, we can calculate the mean value of k and its standard deviation; the reader should carry through the steps and convince himself of the result:

$$\langle k \rangle = \sum_{k=0}^{M} k\ p_k$$

$$= p\ M, \qquad (2.28)$$

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$$

$$= p(1-p)\ M. \qquad (2.29)$$

Hence $$\sigma/\langle k \rangle = \frac{\sqrt{p(1-p)M}}{p\ M} = \sqrt{\frac{(1-p)}{p\ M}}. \qquad (2.30)$$

Again the characteristic inverse square root of the number of molecules appears. Notice that if p is much less than unity - if, for example, we are studying the probability of finding k molecules in a very small volume of the original box - then σ is approximately equal to $\sqrt{p\ M}$, or to $\sqrt{\langle k \rangle}$, and the ratio $\sigma/\langle k \rangle$ is then $1/\sqrt{p\ M}$, or $1/\sqrt{\langle k \rangle}$; the expected fluctuations from the mean number are of the order of the square root of the mean number itself.

2.5 THE MULTINOMIAL DISTRIBUTION

The binomial distribution occurs whenever a choice of two alternatives is made repeatedly. But often there are more than two alternatives! We can imagine dividing our box of gas into three, or ten, or a million parts instead of two; a random walk can take place on a plane or in space instead of on a line, with a number of different choices of steps possible each time rather than simply "right" or "left"; a magnetic atom may have several orientations possible relative to an external magnetic field instead of simply "along" or "against." How are we to handle these cases?

We may state the problem this way: M molecules are to be distributed among N cells of equal volume into which the box has been divided in imagination. How many of the possible assignments of individual molecules to cells correspond to having k_1 molecules in cell 1, k_2 in cell 2, and so on up to k_N in cell N, in such a way that $k_1 + k_2 + \ldots + k_N = M$? Let us first notice that our previous work has shown that for $N = 2$, the number in question is $M!/(k_1!)(k_2!)$, using the new notation. We can understand this result by slightly different reasoning than we originally used to obtain it. The number of different assignments would be

M! if we had only one molecule in each cell, since M! is the number of permutations of M objects. But a permutation of the k_1 molecules in the first cell among themselves, when k_1 is greater than one, does not give a distinct assignment, so we must divide by $k_1!$; a similar argument for the other cell requires us to divide by $k_2!$. Now the required generalization to N cells is easy to see: We must divide by the factorial of each individual cell's population. Hence the result for the required number of assignments, which we shall call $n(k_1, k_2, \ldots k_N)$ is:

$$n(k_1, k_2, \ldots k_N) = \frac{M!}{k_1!\, k_2! \ldots k_N!}. \qquad (2.31)$$

The total number of assignments is N^M, since there are N possibilities for each of the M molecules; hence the probability of the given assignment is

$$p(k_1, k_2, \ldots k_N) = \frac{M!}{N^M\, k_1!\, k_2! \ldots k_N!}. \qquad (2.32)$$

This distribution, which is a joint distribution of the kind discussed in section 1.4 in N variables, is known as the multinomial distribution. As with the binomial distribution, this name originates from the fact that the $n(k_1, k_2, \ldots k_N)$ are the coefficients in a certain expansion:

$$(a_1 + a_2 + \ldots a_N)^M = \sum_{(k_i)} n(k_1, k_2, \ldots k_N)(a_1)^{k_1}(a_2)^{k_2} \ldots (a_N)^{k_N}, \qquad (2.33)$$

where the sum is over all sets of non-negative integers $k_1, k_2, \ldots k_N$ satisfying the condition $k_1 + k_2 + \ldots k_N = M$. This is the general multinomial expansion; that the coefficients are indeed given by Eq. (2.31), follows on observing that a given coefficient counts the number of distinct ways of picking out k_1 factors a_1, k_2 factors a_2, ... k_N factors a_N from M factors $(a_1 + a_2 + \ldots a_N)$. This is exactly the same counting problem as that of picking out k_1 molecules to put on the first cell, etc., out of a total of M molecules. We can use this to confirm that the total number of assignments is N^M; indeed, setting each of the a_i in Eq. (2.33) equal to unity, we get:

$$\underset{\text{N terms}}{(1 + 1 + \ldots 1)^M} = \sum_{(k_i)} n(k_1, k_2, \ldots k_N)(1)^{k_1}(1)^{k_2} \ldots (1)^{k_N}$$

or

$$N^M = \sum_{(k_i)} n(k_1, k_2, \ldots k_N); \qquad (2.34)$$

(k_1,k_2,k_3)	$n(k_1,k_2,k_3)$	PERMUTATIONS	PRODUCT
(5,0,0)	1	3	3
(4,1,0)	5	6	30
(3,2,0)	10	6	60
(3,1,1)	20	3	60
(2,2,1)	30	3	90
			243 = 3

Table 2.3

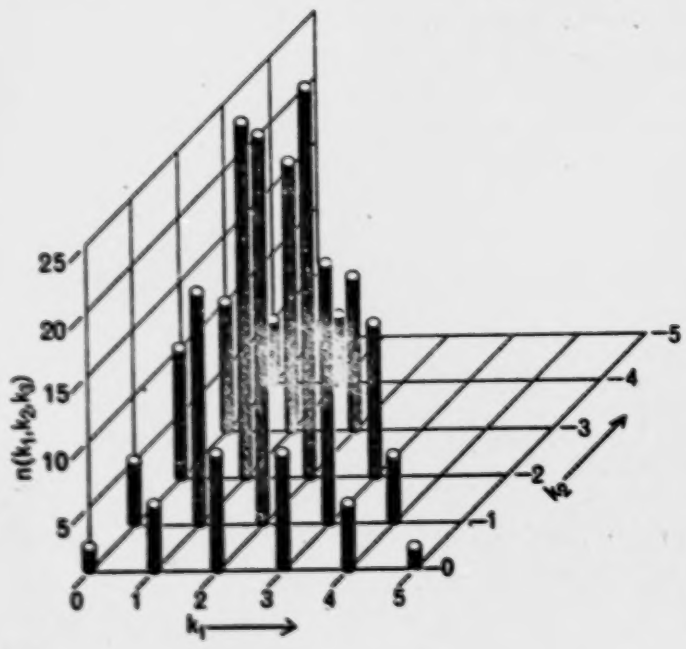


Fig. 2.6

hence the probabilities add correctly to unity:

$$\sum_{(k_i)} p(k_1, k_2, \ldots k_N) = \sum_{(k_i)} \frac{1}{N^M} n(k_1, k_2, \ldots k_N) = \frac{1}{N^M} \cdot N^M = 1. \qquad (2.35)$$

This distribution is harder to visualize than the binomial. For N = 3, however, where k_1 and k_2 are essentially the only variables involved (since k_3 is necessarily equal to $(M - k_1 - k_2)$, we can construct a three-dimensional graph analogous to Fig. 1.4 of Chapter 1. We have done this in Fig. 2.6, (preceding page) which exhibits the frequencies $n(k_1,k_2,k_3)$ for the case M = 5. Table 2.3 (preceding page) gives the values of n for the possible assignments.

From Fig. 2.6, one can see a tendency, even though the number of molecules is small, for the assignments corresponding to approximately equal distribution of the molecules among the cells to predominate. It is not hard to demonstrate that the uniform (i.e., equal population of cells) distribution is indeed the most probable one, at least in the case where the number M of molecules is an integer multiple of the number N of cells. For if M = rN, where r is an integer, the uniform distribution is the one for which $k_1 = k_2 = \ldots = k_N = r$; the corresponding n is $(M!)/(r!)^N$. Suppose we move <u>one</u> molecule out of one cell and into another; the corresponding n is now $(M!)/(r-1)! \times (r+1)!(r!)^{N-2}$. The ratio of the second to the first is $(r!)^2/(r-1)! \times (r+1)!$, which equals $r/(r+1)$; hence the change has <u>decreased</u> the number of assignments, and thus also the probability of occurrence.

We can calculate the mean value of the k_i's, and also other averages of interest; but for this purpose we must first note the generalization of the definition of an average which is required for the case of a joint distribution. The average of any function of the k_i's is defined in analogy to Eq. (2.7) of section 2.1:

$$\langle f(k_1, k_2, \ldots k_N) \rangle = \frac{1}{N^M} \sum_{(k_i)} f(k_1, k_2, \ldots k_N)\, n(k_1, k_2, \ldots k_N) = \sum_{(k_i)} f(k_1, k_2, \ldots k_N)\, p(k_1, k_2, \ldots k_N), \qquad (2.36)$$

where the sum is over all values of each k_i from 0 to M, such that $k_1 + k_2 + \ldots + k_N = M$. In particular, the mean value of any of the k_i can be calculated for the multinomial distribution given by Eq. (2.31) or (2.32):

$$\langle k_1 \rangle = \frac{1}{N^M} \sum_{(k_i)} \frac{k_1\, M!}{k_1!k_2! \ldots k_N!} = \frac{M}{N^M} \sum_{(k_i)} \frac{(M-1)!}{(k_1 - 1)!\, k_2! \ldots k_N!}$$

and using Eq. (2.34) with (M − 1) in place of M, we obtain:

$$\langle k_1 \rangle = \frac{M}{N^M} \cdot N^{M-1} = \frac{M}{N} \qquad (2.37)$$

As we expect, the mean value $\langle k_1 \rangle$ is simply the number of molecules per cell; an identical calculation holds for each of the k's, from the symmetry of the multinomial distribution in the k's. Hence we have for each i:

$$\langle k_i \rangle = M/N. \qquad (2.38)$$

Similarly, we can find the standard deviation for each k_i; the calculation is like that for the binomial distribution, and yields, for each i:

$$\sigma_i^2 = \langle k_i^2 \rangle - \langle k_i \rangle^2 = (N-1)\, M/N^2. \qquad (2.39)$$

Notice that this agrees with the result (Eq. (2.24)) for the case N = 2. If on the other hand we allow N to be

much greater than unity, this result becomes approximately:

$$\sigma_i^2 = M/N = \langle k_i \rangle, \qquad (2.40)$$

or $$\sigma_i = \sqrt{M/N} = \sqrt{\langle k_i \rangle},$$

and $$\sigma_i/\langle k_i \rangle = 1/\sqrt{M/N} = 1/\sqrt{\langle k_i \rangle}. \qquad (2.41)$$

Thus the relative fluctuations in number in a given cell are again inversely proportional to the square root of the mean number in that cell. With one cubic centimeter of gas containing 10^{20} molecules, for example, the number of molecules in each cubic micron of the volume (one micron = 10^{-4} cm) is 10^8 on the average; this would exhibit fluctuations of the order of 10^4 molecules, which is only 0.01% of the average number.

Again, here, as with the symmetric binomial distribution in section (2.3), we have implicitly assumed, in giving equal weights to all the individual assignments of molecules to cells, that a single molecule is "equally likely" to be found in each cell. This is surely plausible when the volumes of the cells are equal; but what if they are not? Or, in the random walk problem, what if steps in different directions have different probabilities? The generalization required to handle these problems follows along the same lines as given in section 2.4. Suppose the volume of the ith cell is V_i and the total volume is V; then the probability p_i of finding a single molecule in the ith cell is $p_i = V_i/V$; the sum of the p_i's is 1, since the sum of the V_i's is V. The probability of finding k_1 specified molecules in V_1, k_2 in V_2, and so on, is $(p_1)^{k_1}(p_2)^{k_2} \ldots (p_N)^{k_N}$; this must be multiplied by the number of assignments, from Eq. (2.31), that give the required distribution among the cells, to yield the result:

$$p(k_1, k_2, \ldots k_N) = \frac{M!}{k_1!\, k_2! \ldots k_N!} \times (p_1)^{k_1}(p_2)^{k_2} \ldots (p_N)^{k_N}. \qquad (2.42)$$

The multinomial theorem, Eq. (2.33) assures us that

$$\sum_{(k_i)} p(k_1, k_2, \ldots k_N) = 1,$$

and the previous case, given by Eq. (2.32) is recovered when all the cell volumes and hence all the probabilities are equal: $p_1 = p_2 = \ldots = p_N = 1/N$.

The discussion of the mean values and the standard deviations of the k_i's follows very much as before; carrying out the details is left as one of the problems at the end of the chapter. The results are:

$$\langle k_i \rangle = p_i M \qquad (2.43)$$

$$\sigma_i^2 = p_i (1 - p_i) M \qquad (2.44)$$

$$\sigma_i/\langle k_i \rangle = \sqrt{(1 - p_i)/p_i M} \qquad (2.45)$$

These results clearly reduce to the previous ones (Eqs. (2.38) and (2.39)), when $p_i = 1/N$. We see that the mean fraction of the molecules in the ith cell is just p_i, which is simply the fraction of the total volume in the ith cell, and that again the fractional deviation from the mean is inversely proportional to the square root of the mean number for each cell.

Let us summarize in more general language the essential result of this section. Suppose N alternative outcomes of an event are possible, and they have probabilities p_1, p_2, ... p_N. Then if a sequence of M independent trials are carried out, the probability that outcome 1 occurs k_1 times, 2 occurs k_2 times, and so on, is given by the multinomial distribution, Eq. (2.42). In the limit as M becomes larger and larger, the fraction k_1/M is increasingly likely to be found very near to p_1; that is, its mean is p_1, and its standard deviation decreases as M increases. Indeed, it is precisely this behavior of repeated trials which allows us to identify the probability p_i with the frequency of occurrence of the ith alternative in repeated trials as discussed in section 1.6, and which therefore renders consistent our very use of the term.

PROBLEMS

2.1 Work out the distribution for the total showing face up when a pair of dice are thrown, assuming both the dice and the throw to be unbiased (i.e., each of the six faces equally likely to turn up). What is the most probable result? The mean result? The standard deviation?

2.2 Use the binomial distribution to discuss the random walk on a line. Starting at $x = 0$, a particle moves unit distance either to the right or the left with equal probability. (Imagine tossing a coin each move to decide which way - heads, go right, tails, go left.) Out of a total number N of moves, it takes L to the left and R to the right; $L + R = N$. The net distance traveled to the right is evidently $D = R - L$. What is the distribution of D? Find the most probable, and the mean values of D. What is the root mean square of D?

2.3 Coin-tossing sequences of heads and tails may be discussed with the help of the binomial distribution. What odds should you be willing to offer against tossing exactly 6 heads in 20 tosses?

2.4 Toss a coin 60 times, and keep a record of the results.

(a) Does the total number of heads in the 60 throws lie within the theoretical standard deviation of the expected mean number?

(b) Divide the results into 20 sequences of 3 tosses, and find the distribution among the four alternatives (0,1,2,3, heads). Does it lie "reasonably" close to the theoretical binomial distribution?

2.5 For the case of three molecules in a box, work out the results of Table 2.3 in the following way: Imagine first that the box is divided into three equal parts, and list explicitly (as in Fig. 2.4) all of the possible ways of assigning the three molecules to the three parts. Then imagine one of the partitions eliminated, so that there are only two parts left, one double the size of the other. Count the number of assignments corresponding to 0, 1, 2, 3, molecules, respectively, in the larger part.

2.6 Consider a random walk in the plane along a square network with equal possibilities of moving right, left, up, or down at each step. Let M steps be taken, of which k_1, k_2, k_3 and k_4 are respectively to the right, left, up, and down.

(a) Express the distance D from the starting point in terms of the k_i's.

(b) Find the mean square distance moved in M steps, $\langle D^2 \rangle$.

(Hint: The k_i are distributed according to a multinomial distribution with $N = 4$. You will need to work out averages of the form $\langle k_1 k_2 \rangle$ as well as those done in the text.)

2.7 Given a joint distribution of two or more quantities k_1, k_2, ... , the correlation coefficient r_{ij} is defined by the relation

$$r_{ij} = (\langle k_i k_j \rangle - \langle k_i \rangle \langle k_j \rangle)/\sigma_i \sigma_j .$$

Work out the value of r_{ij} for the multinomial distribution. To what does it reduce for the case $N = 2$?

3 CONTINUOUS DISTRIBUTIONS

3.1 INTRODUCTION: MEAN VALUES; EXAMPLES

The probability distributions dealt with in Chapter 2 were <u>discrete</u>, that is, the possible alternatives could be characterized by an integer k (or a set of integers k_1), which ranged over a finite number of values. Very often we need to deal with what we will call <u>continuous</u> distributions, where the possible alternatives are characterized by one or more variables which range over a continuum like that of the real numbers. Some examples of continuously distributed quantities which arise in physics, together with the range of possible values of the quantity in question, are:

(1) The distance between successive collisions of a molecule in a gas - the so-called "free path"; any positive value.

(2) A velocity component of a molecule in a gas; any positive or negative value.

(3) The angle through which a nuclear particle is scattered in a collision; any value between 0 and π.

(4) The time of decay of a radioactive nucleus; any positive value.

Of course, in any of these examples, we may - as in the example of the velocities in Chapter 1 - divide the range into a finite number of intervals, and treat the distribution as discrete. Indeed, the limitations of accuracy of the measuring instrument, and the finiteness of the population may require us to do so. Nevertheless, we may imagine both the accuracy of measurement and the size of the population increased sufficiently to allow the intervals to be decreased indefinitely; in the limit we can speak about the probability corresponding to an arbitrarily small interval.

It is simplest to begin with the notion of the <u>cumulative</u> distribution function, which was discussed in section 1.3. Let us denote by P(x) the probability that the quantity whose distribution is under study is less than x. Then P(x) can never decrease as x increases; for if x' is greater than x, P(x') must be equal to P(x), the probability that the quantity is less than x, plus the probability that it lies between x and x'; since probabilities cannot be negative, P(x') cannot be less than P(x). The general appearance of possible P's is illustrated in Figs. 3.1 and 3.2. The first shows a distribution which has a minimum and a maximum possible value for x, while the second shows a distribution which extends indefinitely to large negative and positive values. Ultimately, P(x) must approach 0 at the left-hand end and 1 at the right-hand end of the graph. The probability of finding a value of x lying <u>between</u> two specified values x_1 and x_2 is then given (provided x_1 is less than x_2) by $[P(x_2) - P(x_1)]$. If P(x) is a continuous function, then $[P(x_2) - P(x_1)]$ will ap-

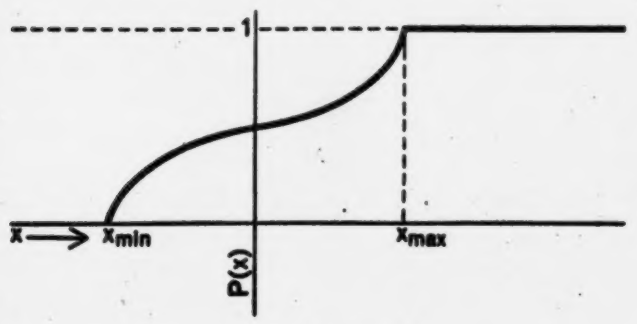


Fig. 3.1

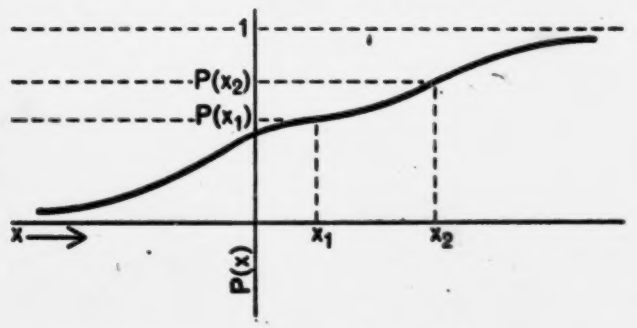


Fig. 3.2

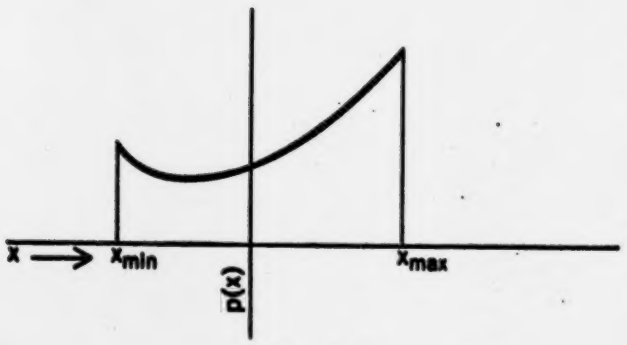


Fig. 3.3

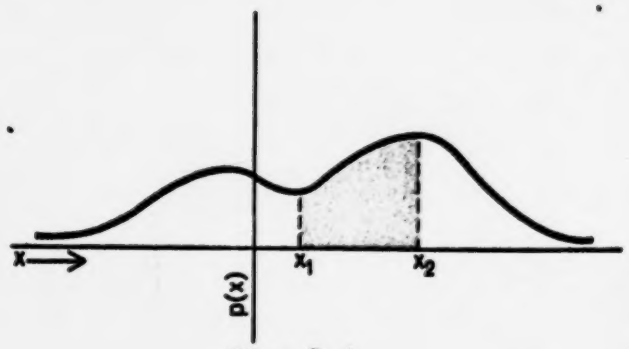


Fig. 3.4

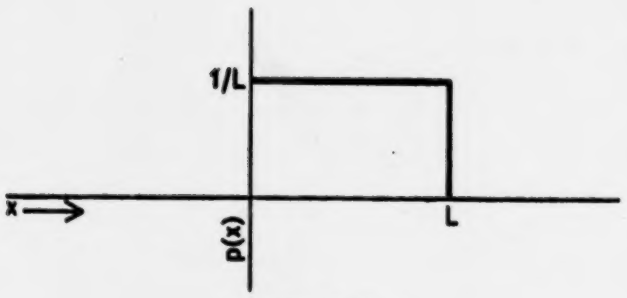


Fig. 3.5

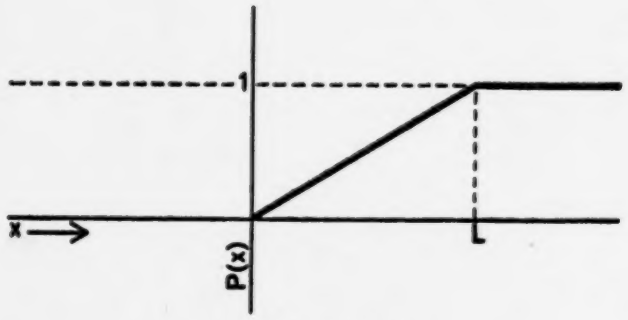


Fig. 3.6

proach zero as x_2 approaches x_1; and if furthermore, P(x) is differentiable, then for $(x_2 - x_1)$ sufficiently small, we may write by Taylor's expansion:

$$P(x_2) = P(x_1) + (x_2 - x_1)(dP/dx)_{x=x_1},$$

neglecting higher terms of the expansion.

Hence

$$P(x_2) - P(x_1) = (x_2 - x_1)\, p(x_1), \quad (3.1)$$

where $$p(x) = dP/dx. \quad (3.2)$$

Thus the probability that x lies in a small interval around x_1 is given by the product of the interval with $p(x_1)$; hence the name probability density function which is sometimes used for the derivative p(x). It can also be called the probability "per unit interval of x." Dimensionally it has the units of the reciprocal of x, since multiplication by an interval of x gives a pure number - a probability. Note particularly that it is not the probability of "finding the value x"; since we are dealing with a continuum, such a probability must be zero! Figures 3.3 and 3.4 illustrate the probability density functions corresponding to the cumulative probability functions of Figs. 3.1 and 3.2, respectively. The cross-hatched area in Fig. 3.4 is given by the integral of p(x) from x_1 to x_2. Using Eq. (3.2) and the fundamental theorem of integral calculus, we have:

$$\int_{x_1}^{x_2} p(x)\, dx = \int_{x_1}^{x_2} (dP/dx)\, dx$$

$$= P(x)\Big|_{x_1}^{x_2} = P(x_2) - P(x_1). \quad (3.3)$$

Hence the area under the probability density curve from x_1 to x_2 gives the probability of finding a value of x lying between x_1 and x_2. In particular, if we let x_1 go to $-\infty$, we have, since $P(-\infty) = 0$:

$$\int_{-\infty}^{x_2} p(x)\, dx = P(x_2), \quad (3.4)$$

and if we let x_2 go to $+\infty$, we have, since $P(+\infty) = 1$:

$$\int_{-\infty}^{\infty} p(x)\, dx = 1 \quad (3.5)$$

The equation is completely analogous to Eq. (2.3) of the discrete case, and is sometimes referred to as the "normalization" condition on the probability density function.

We may define mean values in a fashion analogous to that use in Chapter 2, but replacing sums by integrals; thus the mean value of x is:

$$\langle x \rangle = \int_{-\infty}^{\infty} x\, p(x)\, dx, \qquad (3.6)$$

and the standard deviation is given by

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle$$
$$= \int_{-\infty}^{\infty} (x - \langle x \rangle)^2\, p(x)\, dx$$
$$= \langle x^2 \rangle - \langle x \rangle^2 \qquad (3.7)$$

and in general, the mean value of any function of x by

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x)\, p(x)\, dx \qquad (3.8)$$

These formulae may be thought of as arising from the corresponding ones of the discrete case (Eqs. (2.5), (2.7), and (2.11)) by first breaking up the range of x into N intervals, and associating the probability $p(x_k)\Delta x_k$ with the kth interval Δx_k. Applying the discrete formulae, and then passing to the limit $\Delta x_k \to 0$ and $N \to \infty$, gives the integral expressions.

Let us illustrate these notions with a few examples.

3.1.1 Uniform Distribution Between x = 0 and x = L

Suppose a particle is "equally likely" to be found anywhere on the x-axis between 0 and L. We take this to mean that the probability density p(x) is constant between 0 and L, and zero elsewhere:

$$p(x) = \begin{cases} 0, & (x < 0) \\ C, & (0 \le x \le L) \\ 0, & (x > L) \end{cases} \qquad (3.9)$$

Since condition (3.5) must be satisfied, the constant C must be equal to 1/L. The cumulative distribution P(x) is then given by application of Eq. (3.4):

$$P(x) = \int_{-\infty}^{x} p(x')\, dx'$$

$$= \begin{cases} 0, & (x < 0) \\ x/L, & (0 \le x \le L). \\ 1, & (x > L) \end{cases} \qquad (3.10)$$

Figures 3.5 and 3.6 show graphs of p(x) and P(x), respectively. A simple calculation, using Eqs. (3.6) and (3.7), shows that $\langle x \rangle = L/2$, as one would expect, and that $\sigma = L/\sqrt{12} = 0.289\ L$.

3.1.2 Uniform Distribution in Angle

Imagine a particle scattered so that its final direction of travel makes an angle θ with its initial direction. We say it has been scattered through an angle θ. If we are considering scattering only in a plane, θ may range from $-\pi$ to $+\pi$, and a uniform distribution would correspond to a density function $p(\theta) = 1/2\pi$, in analogy with the result in 3.1.1. Suppose, however, that we are interested in scattering in space, and that by uniformity of distribution we mean that all directions in space, relative to the original direction, are "equally likely." A reasonable interpretation of what this means is the following: Consider a sphere; each point of the sphere determines by its radius vector from the center a direction of scattering, relative to a fixed direction determined

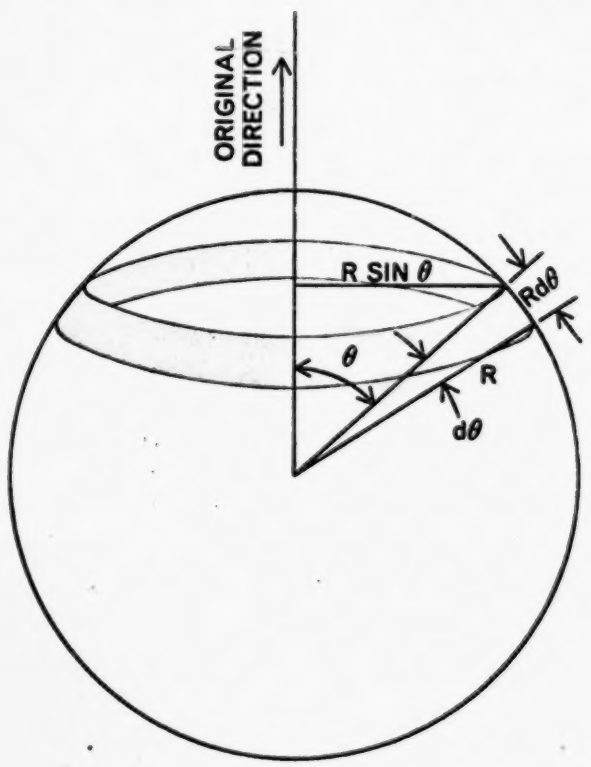


Fig. 3.7

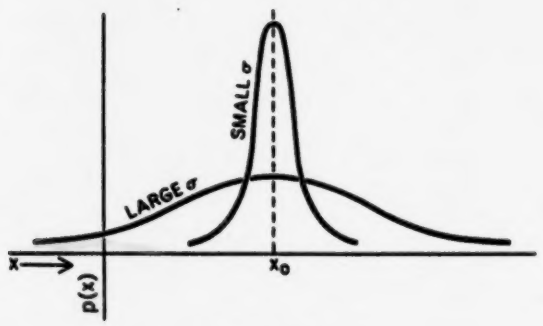


Fig. 3.8

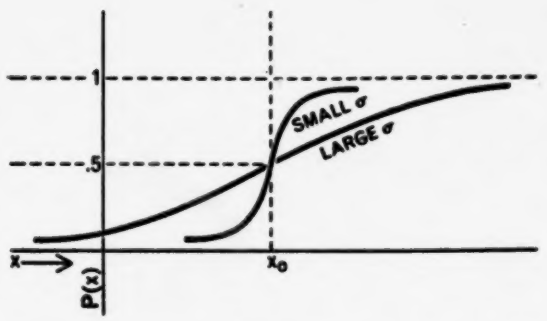


Fig. 3.9

by a fixed point on the sphere. Then we shall mean by "all directions equally likely," that equal areas on the sphere are equally likely; hence the probability of scattering through an angle between θ and $(\theta + d\theta)$, relative to a given direction, is to be equated to the ratio of the area of the sphere corresponding to such directions, to the whole area of the sphere. Referring to Fig. 3.7, (preceding page) we see that angles of scattering between θ and $(\theta + d\theta)$ correspond to a ring of radius $R \sin\theta$ and width $R\, d\theta$ on the sphere; the area of the ring is therefore $2\pi R^2 \sin\theta\, d\theta$, and since the surface area of the whole sphere is $4\pi R^2$, we have:

$$p(\theta)\, d\theta = 2\pi R^2 \sin\theta\, d\theta / 4\pi R^2$$

$$= \tfrac{1}{2} \sin\theta\, d\theta, \quad (0 \le \theta \le \pi). \tag{3.11}$$

Thus the probability density vanishes at both 0 and π, and has a maximum at $\pi/2$, that is, it vanishes for directly forward or directly backward scattering, and has a maximum at right angles to the incident direction. There are more ways, so to speak, in which scatterings at right angles can occur, than forward and backward scatterings.

3.1.3 Gaussian Distribution

This distribution occurs frequently in a variety of applications. We shall see it arising as a limiting case of the binomial distribution, and in the discussion of the distribution of errors of observation. It is also called the "normal distribution" by statisticians. It is defined by:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -(x-x_0)^2/2\sigma^2, \tag{3.12}$$

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp -(x'-x_0)^2/2\sigma^2\, dx'.$$

The mean value of the distribution is x_0 and the standard deviation is σ. Figures 3.8 and 3.9, respectively, give graphs of $p(x)$ and $P(x)$; in each one, two cases are included, corresponding to "small" and "large" values of σ. The integral expression for $P(x)$ in Eq. (3.13) cannot be expressed in terms of "elementary" functions, but itself defines a new function, values of which can be found in mathematical or statistical tables.

3.2 CHANGE OF VARIABLE

A frequently arising problem in dealing with distributions is that of finding the distribution of a quantity which is a known function of another quantity, whose distribution is known. Consider, for example, the angular distribution of particles scattered in collisions. What determines its form? If the scattering can be treated by classical dynamics, the angle of scattering in any given collision is determined by the "impact parameter" of the collision, which is the perpendicular distance from the line along which the incident particle travels, to a parallel line through the scattering center. That is, there is a unique functional relation, whose form de-

pends on the nature of the interaction between the incident particle and the scattering center, between the impact parameter b and the scattering angle θ. Figure 3.10 illustrates a possible pair of trajectories for a scattering event, with the impact parameter angle indicated for each. To determine the distribution of scattering angles, therefore, one must know the distribution of impact parameters, and translate through the functional relation between the two.

To state the problem in general terms, suppose we are given the probability density function p(y) for a quantity y, and we are also given that y is related to another quantity x through a functional relation $y = f(x)$. What is the probability density function for x? We will assume for simplicity initially that f(x) is monotone; that is, that it is a steadily increasing function. The reason for this is to assure that each value of x corresponds to only one value of y and vice versa. Figure 3.11 illustrates the relation between y and x. Referring to the figure, we see that the values of y lying between y_0 and $(y_0 + \Delta y)$ correspond precisely to the values of x lying between x_0 and $(x_0 + \Delta x)$ and to no others, where $y_0 = f(x_0)$, and Δy is related to Δx through:

$$\Delta y = f'(x_0)\Delta x, \quad f'(x) = df/dx. \tag{3.14}$$

The probability that y lies between y_0 and $(y_0 + \Delta y)$ is $p(y_0)\Delta y$; this is identical with the probability that x lies in its corresponding interval. Hence we have:

$$\text{probability x is in interval from } x_0 \text{ to } (x_0 + \Delta x) = p(y_0)\Delta y = p[f(x_0)]f'(x_0)\Delta x. \tag{3.15}$$

This is in the form of a function of x_0 multiplied by the interval Δx; hence from the definition of the probability density function for x, the function of x_0 multiplying Δx must be that density. Denoting it by q(x), we have:

$$q(x) = p[f(x)]f'(x). \tag{3.16}$$

We may say: substitute f(x) for y both in the density and in the differential of y; the coefficient of the differential of x is the density function for x. Suppose now that f(x) had been a decreasing function instead of an increasing one. The slope f'(x) would have been negative, and if we used Eq. (3.16) unchanged, we would have a negative probability, which is not allowed. The difficulty can be traced to Eq. (3.14), which gives the relation between the intervals Δx and Δy. We really only want the relation between their magnitudes; for our purposes it is irrelevant whether the slope is positive or negative. We can take this into account by using the absolute magnitude of the slope in Eq. (3.14); we then have:

$$\Delta y = |f'(x)|\Delta x \tag{3.17}$$

and Eq. (3.16) would become instead:

$$q(x) = p[f(x)]|f'(x)|. \tag{3.18}$$

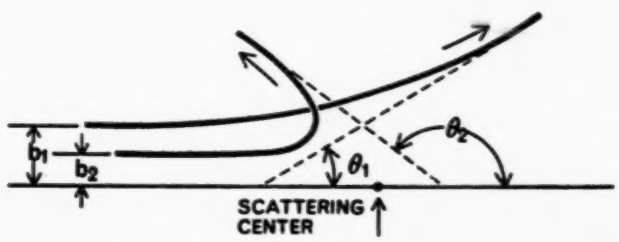


Fig. 3.10

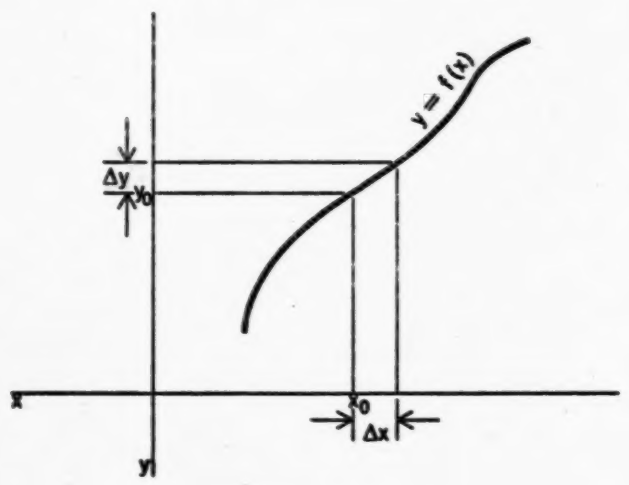


Fig. 3.11

If now we consider a case in which y is <u>not</u> a single-valued function of x, that is, more than one value of y corresponds to certain values of x, the probability density for x will have contributions from <u>each</u> y corresponding to that x. Figure 3.12 shows a possible situation of this kind. From the figure one sees that for x between x_1 and x_2, there is one contribution to q(x), whereas between x_2 and x_3 there are two contributions. If we distinguish between upper and lower branches of f(x) by calling them $f_1(x)$ and $f_2(x)$, as indicated in Fig. 3.12, then the probability density function for x is given by:

$$q(x) = \begin{cases} p[f_1(x)]f_1'(x), & (x_1 < x < x_2) \\ p[f_1(x)]f_1'(x) + p[f_2(x)] \\ \qquad \times |f_2'(x)|, & (x_2 < x < x_3). \end{cases} \tag{3.19}$$

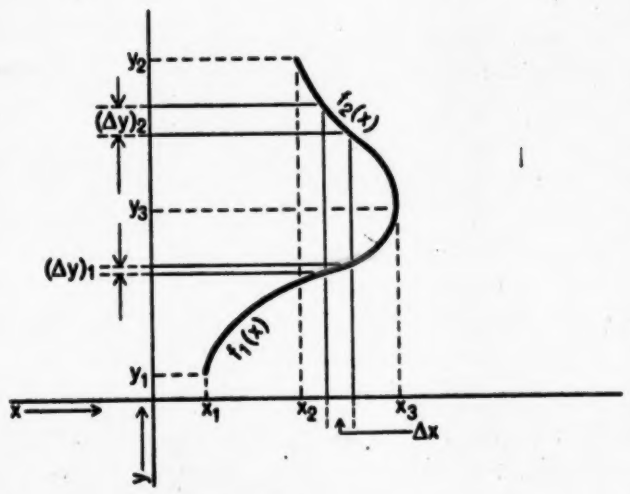


Fig. 3.12

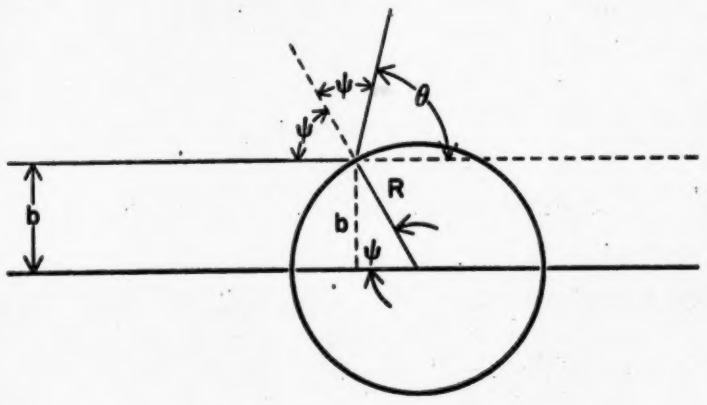


Fig. 3.13

More complicated cases can be treated similarly.

Let us give an instance of each of these possibilities. The scattering problem can be illustrated by a very simple model: Imagine particles scattered by a smooth sphere. The relation between impact parameter and scattering angle is easily found, using the geometry of the problem. Figure 3.13 shows a typical trajectory; the particle is incident at impact parameter b, bounces off the smooth sphere of radius R with angle of "reflection" equal to angle of incidence, and goes off at angle θ with its original direction. From the figure, we see that the following relations hold:

$$\theta + 2\psi = \pi,$$

and $$\sin \psi = b/R.$$

Hence we have for the relation between b and θ:

$$b = R \sin \psi = R \sin \left[\tfrac{1}{2}(\pi - \theta)\right]$$
$$= R \cos (\theta/2) \tag{3.20}$$

We are now ready to translate the b distribution to θ distribution. But what is a reasonable assumption about the distribution of impact parameters? If we are thinking of the problem as a model for an atomic or nuclear scattering experiment, we are not able to control the b values for individual collisions; we simply have a beam of particles of a certain mean intensity. This means that, in a plane perpendicular to the beam, equal areas have, on the average, equal numbers of particles incident on them. That is, the probability density is uniform across the area of the beam. Relative to a given target particle (the sphere of our model), this means that the probability of the impact parameter lying between b and (b + db) is given by the ratio of the area perpendicular to the beam corresponding to this range of b values, to the total area of the beam intercepted by the sphere. (In this

model, particles incident at values of b greater than R are not scattered at all, and will not enter into the discussion.) Since the area of a ring of radius b and width db is $2\pi b\ db$, and the area of a circle of radius R is πR^2, we have:

$$p(b)\ db = 2\pi b\ db/\pi R^2$$

$$= 2\ b\ db/R^2, \qquad (0 \leq b \leq R). \tag{3.21}$$

We can now apply Eq. (3.16), to find the distribution of θ:

$$q(\theta) = 2R \cos\left(\frac{\theta}{2}\right)\left|\frac{d}{d\theta}\left[R \cos\left(\frac{\theta}{2}\right)\right]\right|/R^2$$

$$= 2 \cos(\theta/2)\ \tfrac{1}{2} \sin(\theta/2)$$

$$= \tfrac{1}{2} \sin\theta \qquad (0 \leq \theta \leq \pi). \tag{3.22}$$

Remembering the discussion of the uniform angular distribution in section 3.1.2, we see that the distribution above is identical with that found in Eq. (3.11) of that section; the angular distribution of the particles scattered from a smooth sphere is uniform. This simple result is, so to speak, an accident arising from the particular form of interaction between the incident particle and the scattering center. Other laws of interaction will give different angular distributions; in particular, the Coulomb interaction which enters when the particles both carry electrostatic charges is an important case, but will be left to the Problems for discussion.

For an illustration of the case of a multiple-valued relation between the quantities in question, consider the following example. A simple harmonic oscillator is vibrating with amplitude a and angular frequency ω, so that its displacement x at time t is given by:

$$x = a \sin \omega t \tag{3.23}$$

The period T is related to ω through $\omega T = 2\pi$.

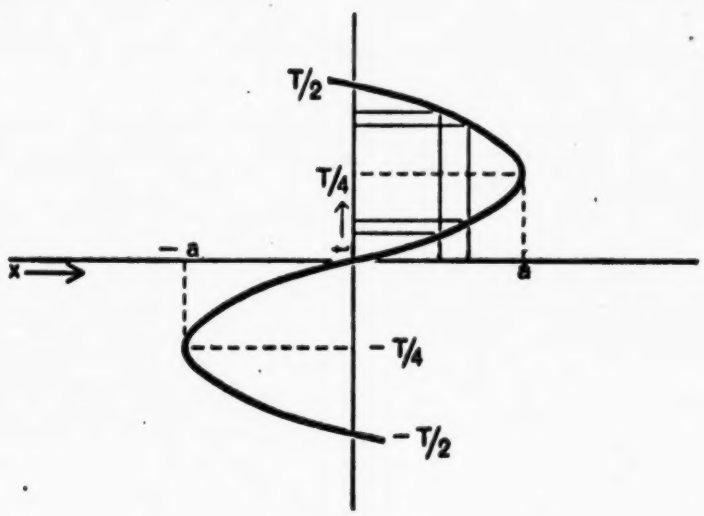


Fig. 3.14

Suppose we look at, or take snapshots of the oscillator "at random." What is the distribution of displacements we will observe? By "randomness" here, we mean that any instant is as likely to be chosen as any other for the snapshot; that is, the distribution of observation <u>times</u> is uniform, in the same sense as we used the term in section 3.1.1. From the uniformity of distribution of observation times, we are to deduce the corresponding distribution of displacements, given Eq. (3.23) relating the two. Figure 3.14 shows the relation between t and x. It is many-valued, but we may confine our attention to a single period of the motion, since all periods are identical. Within one period, say from $-T/2$ to $+T/2$, there are two time intervals which correspond to any one given space interval, as shown in the figure. However, the symmetry of the curve is such that they will each contribute equally to the distribution of x, so that we can confine our attention to the half period from $-T/4$ to $+T/4$, during which x increases monotonically from $-a$ to $+a$. A uniform distribution of times in this interval corresponds to the density function:

$$p(t)\ dt = (2/T)\ dt, \quad (-T/4 \leq t \leq +T/4). \tag{3.24}$$

The relation between t and x is found

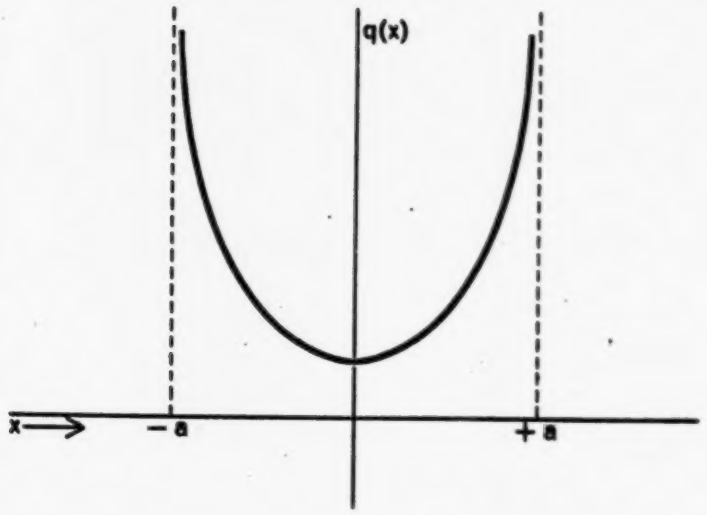


Fig. 3.15

by solving Eq. (3.23) for t:

$$t = (1/\omega) \arcsin (x/a). \qquad (3.25)$$

Hence we have, applying Eq. (3.14):

$$q(x) = \left(\frac{2}{T}\right) \frac{d}{dx}\left[\left(\frac{1}{\omega}\right) \arcsin \left(\frac{x}{a}\right)\right]$$

$$= \left(\frac{2}{\omega T}\right) \frac{1}{\sqrt{1 - (x/a)^2}} \cdot \frac{1}{a}$$

$$= \frac{1}{\pi\sqrt{a^2 - x^2}} \qquad (-a \le x \le +a). \qquad (3.26)$$

This probability density function for x is sketched in Fig. 3.15. We see that it has a symmetrical minimum at x = 0, and that it rises asymptotically to ∞ at x = −a and x = +a. This is no cause for alarm, if we remember that it is only areas under the curve that are interpreted as probabilities. Any interval (even one extending to a or −a) does indeed have a finite area under it; in fact, the area under the whole curve is 1, as it must be. The form of the curve shows that we are much more likely to see large displacements than small ones, if we look at random times; this is understandable when we remember that the speed of the oscillator is large at the center and zero at the end points. It, therefore, spends more time at the ends than at the middle, and our probability density for the observed positions reflects this fact.

3.3 RADIOACTIVE DECAY; MOLECULAR FREE PATHS

An important family of distributions which occurs in various contexts in physics can be illustrated by the problem of radioactive decay. Suppose we observe a substance containing radioactive nuclei, which decay into a different species of nucleus with the emission of an alpha particle, say, at the moment of decay. What is observed is this: The number of nuclei which still survive at time t is an exponential function of time.

$$N(t) = N_0 e^{-\lambda t}. \qquad (3.27)$$

That is, the fraction surviving after any given time is the same for each succeeding time interval of the same length, so that if half remain after 10 days, one quarter will remain after 20 days, one eighth after 30 days, and so on. Stated otherwise, the number which decay between times 0 and t is given by:

$$N_{dec} = N_0 - N(t) = N_0(1 - e^{-\lambda t}) \qquad (3.28)$$

We interpret this in terms of the individual nuclei by saying that the probability of decay between time 0 and time t for each nucleus is:

$$P(t) = 1 - e^{-\lambda t} \qquad (3.29)$$

This is then a cumulative probability distribution for the time of decay; the corresponding probability density function p(t) dt, which gives the probability that the decay occurs between times t and (t + dt), is given as usual by Eq. (3.2):

$$p(t)\,dt = P(t + dt) - P(t)$$

$$= P'(t)\,dt$$

$$= \lambda e^{-\lambda t}\,dt \qquad (3.30)$$

Figs. 3.16, 3.17, 3.18, respectively, give sketches of p(t), P(t), and of $[1 - P(t)]$. Notice that we can interpret Equation (3.30) in the following way: It is the product of two factors; the first, $e^{-\lambda t}$, is, by Eq. (3.29), equal to $[1 - P(t)]$, the probability of survival of time t; the second, λdt, must therefore represent the probability of decay in the interval dt. Now Eq. (3.29) gives, for sufficiently small time intervals Δt (small enough so that $\lambda \Delta t \ll 1$), by expanding the exponential function:

$$P(\Delta t) = 1 - e^{-\lambda(\Delta t)}$$

$$= 1 - [1 - \lambda(\Delta t) \ \ldots \ ]$$

$$= \lambda(\Delta t). \qquad (3.31)$$

The remarkable fact of nature to observe here is that λ is <u>independent of time</u>, that is, the probability of decay in a small time interval Δt is always $\lambda(\Delta t)$, whether we are observing the nucleus immediately after its formation, or after a long time has already passed. The nucleus has, so to say, no memory of its own "age" built in. Contrast this behavior with the distribution of ages in a biological population, like human beings. Here the probability of death in the next year, say, increases steadily from birth on (except possibly in the very early part of life). That is, λ is not constant, but increases with time. A simple model to describe this behavior may be made by assuming λ to be proportional to the time:

$$\lambda(t) = bt. \qquad (3.32)$$

Then we have the probability of death between times t and $(t + dt)$ given by the product of the probability of survival up to time t, that is $[1 - P(t)]$, with the probability of death in the interval dt:

$$p(t)\,dt = [1 - P(t)]\lambda(t)\,dt,$$

or, using Eq. (3.32) and the fact that $p(t) = P'(t)$:

$$P'(t) = bt\,[1 - P(t)],$$

$$d/dt\,[1 - P(t)] = -bt\,[1 - P(t)].$$

This can be integrated, bearing in mind that $P(0) = 0$, to give:

$$1 - P(t) = \exp -(bt^2/2) \qquad (3.33)$$

$$P(t) = 1 - \exp -(bt^2/2) \qquad (3.34)$$

$$p(t) = P'(t) = bt \exp -(bt^2/2). \qquad (3.35)$$

These curves are sketched in Figs. 3.19, 3.20, and 3.21. Notice the similari-

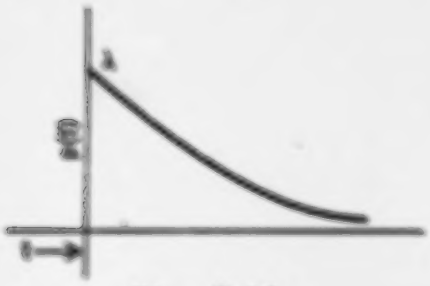


Fig. 3.16

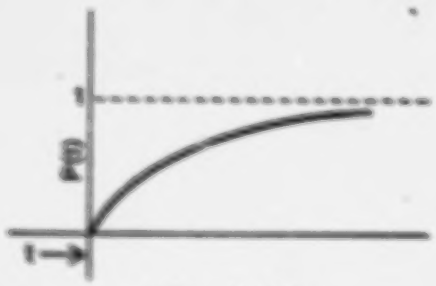


Fig. 3.17

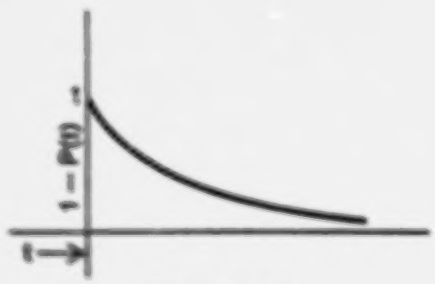


Fig. 3.18

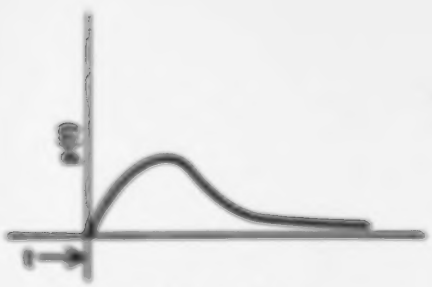


Fig. 3.19

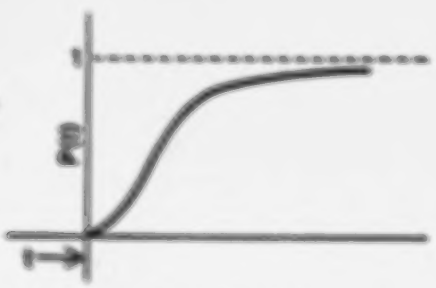


Fig. 3.20

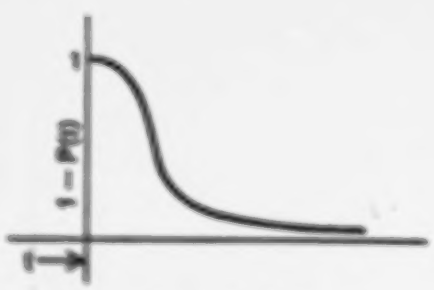


Fig. 3.21

ties and differences between these and the corresponding curves for the case of radioactive decay in Figs. 3.16, 3.17, and 3.18. The difference between the two cases may be put in the following way: for the radioactive decay case, the probability of survival to time $(t_1 + t_2)$ is given by $\exp -\lambda(t_1 + t_2)$; this may be expressed as the product of the probability of survival to time t_1, $\exp -\lambda t_1$, by $\exp -\lambda t_2$, which is the probability of survival a further time t_2, but which is independent of t_1! It is quite otherwise for the survival law given by Eq. (3.33); here we have instead:

$$\exp -b(t_1 + t_2)^2/2 = \exp -bt_1^2/2 \times \exp -b(t_1 t_2 + t_2^2/2),$$

and the second factor, which gives the probability of survival a further time t_2, clearly depends on t_1. This will be the case for any survival law except the radioactive decay law.

Let us go on to calculate the mean life of the decaying nuclei. This will be given by our standard equation for mean values:

$$\langle t \rangle = \int_0^\infty tp(t)\,dt = \int_0^\infty t\lambda e^{-\lambda t}\,dt = (1/\lambda)\int_0^\infty xe^{-x}\,dx$$

Here we have made the change of integration variable $x = \lambda t$. The value of the definite integral occurring above is just 1, so we have:

$$\langle t \rangle = 1/\lambda. \qquad (3.36)$$

Thus the parameter λ characterizing the decay is itself simply the reciprocal of the mean life, and we may re-express the decay law in the form:

$$p(t)dt = \frac{dt}{\langle t \rangle} \exp -t/\langle t \rangle \qquad (3.37)$$

$$P(t) = 1 - \exp -t/\langle t \rangle. \qquad (3.38)$$

The mean life is therefore the time in which the number of surviving nuclei falls to 1/e of its original value.

Essentially the same types of distribution are encountered in a quite different physical context, in which we have "distance of travel" in place of the "time" of the decay law. Of course, a translation from time to distance could be imagined easily for the radioactive decay problem as follows: Imagine all the nuclei to have the same initial speed; then the distance they travel before decaying will be proportional to the time elapsed before decay, and if we can observe the path lengths traveled until decay, they will have a distribution which will be of precisely the same form as our distribution of decay times. (Something very much like this is actually done in fundamental particle physics, where the tracks of unstable short-lived particles are observed in bubble chambers or other devices, and mean lives are deduced from them.) But now we may imagine, instead of "decay," that the particle undergoes "interactions," or more simply, collisions which change its speed and direction of motion. This occurs continually, for example, in a gas; the molecules move in essentially straight lines (so-called "free paths"), interrupted from time to time by collisions which change the direction of motion. Imagine following a molecule in its motion; there will be a sequence of free paths of varying lengths, which therefore have a certain distribution, characterized by a certain mean free path. This distribution is of considerable importance in the kinetic theory of gases, particularly in connection with processes like heat conduction and diffusion for which the collisions play an important role.

We may analyze the problem of the distribution of free paths in much the same way as the radioactive decay problem. The probability that a collision will occur in a small interval Δx is, we assume, simply proportional to the interval, and does not depend on how far the molecule has traveled since its last collision. The molecule has no

"memory" of its past. Thus we will write $(\Delta x/L)$ for this probability, where L is a constant of the dimensions of length. Then if $p(x)\,dx$ is the probability that a collision occurs between x and $(x + dx)$, and if $P(x)$ is the cumulative probability that a collision occurs between 0 and x, so that $[1 - P(x)]$ is the probability that no collision occurs between 0 and x, we must have

$$p(x)\,dx = [1 - P(x)]\,(dx/L). \quad (3.39)$$

But

$$p(x)\,dx = P(x + dx) - P(x) = P'(x)\,dx,$$

hence

$$P'(x) = [1 - P(x)]/L,$$

or

$$d[1 - P(x)]/dx = -(1/L)[1 - P(x)]. \quad (3.40)$$

The solution of this differential equation for which $P(0) = 0$ is:

$$1 - P(x) = \exp -x/L, \quad (3.41)$$

or

$$P(x) = 1 - \exp -x/L, \quad (3.42)$$

and

$$p(x)\,dx = \exp(-x/L)\,dx/L. \quad (3.43)$$

Thus the distribution of path lengths is exponential, exactly like the distribution of radioactive decay times. Figures 3.16, 3.17, and 3.18 serve equally well to illustrate either one.

The mean free path is given in the usual way by:

$$\langle x \rangle = \int_0^\infty xp(x)\,dx$$
$$= \int_0^\infty \exp(-x/L)\,dx/L$$
$$= L. \quad (3.44)$$

Thus the parameter L is simply the mean free path. Its actual value for gases at standard conditions is of the order of 10^{-5} cm.

3.4 OTHER MOLECULAR PATH PROBLEMS; POISSON DISTRIBUTION

We may go on to consider further questions associated with molecular paths such as the following: In a given length of the molecule's path, what is the probability of its undergoing precisely 0,1,2, ... n, . . collisions? What is the probability that the nth collision (from a given starting point) takes place between x and $(x + dx)$? Apart from the intrinsic interest of these questions, their discussion will lead us to a connection with work of the previous chapter, namely to an important limiting case of the binomial distribution.

Consider the first question posed above. Given a length of a molecule's path, what are the respective probabilities that during that length of path the molecule has undergone 0 or 1 or 2 ... collisions? If we denote these probabilities by $Q_0(x)$, $Q_1(x)$, $Q_2(x)$, ... then we already know $Q_0(x)$; since it is the probability of no collision occurring between 0 and x, it is identical with $[1 - P(x)]$ of the previous section, and is given by $\exp -x/L$. What about $Q_1(x)$? There are many ways in which one collision can occur in a path length x; it can occur at any point between 0 and x. The total probability is given by the sum of all the probabilities of the alternative ways in which it can happen. If we denote by dx_1 the small interval in which the collision occurred, and let x_1 be its distance from the origin of the path, we can see with the help of Fig. 3.22, that the probability that one

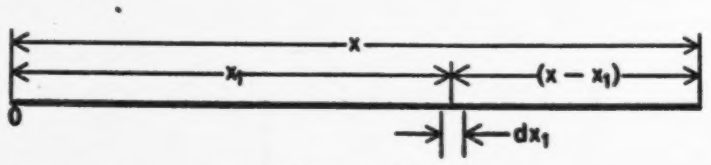


Fig. 3.22

collision occurs in dx_1, and that no collision occurs elsewhere, is given by the product: (probability of no collision between 0 and x_1) times (probability of a collision in dx_1) times (probability of no collision between x_1 and x), that is, by the expression:

$$\exp -\left(\frac{x_1}{L}\right)\frac{dx_1}{L} \quad \exp -\left[\frac{(x-x_1)}{L}\right].$$

This expression must be summed - that is, integrated - over all possible values of x_1, i.e., from 0 to x, to get $Q_1(x)$, the probability of precisely <u>one</u> collision occurring between 0 and x. Hence we have

$$\begin{aligned} Q_1(x) &= \int_0^x \exp\left(-\frac{x_1}{L}-\frac{x-x_1}{L}\right)\frac{dx_1}{L} \\ &= e^{-x/L}\int_0^x \frac{dx_1}{L} \\ &= \left(\frac{x}{L}\right)e^{-x/L}. \end{aligned} \tag{3.45}$$

This function rises from zero at the origin to a maximum value at $x = L$, and then falls again as x increases further. This is to be expected: path lengths of the order of one mean free path are more likely to contain precisely one collision than either longer or shorter paths.

The subsequent functions $Q_2(x)$, $Q_3(x)$ and so on, can be found similarly. With the help of the sketch in Fig. 3.23, we can write the probability that collisions occur in dx_1 at x_1, and in dx_2 at x_2, and nowhere else as:

$$\exp\left(-\frac{x_1}{L}\right)\frac{dx_1}{L} \quad \exp -\frac{(x_2-x_1)}{L} \quad \frac{dx_2}{L}$$

$$\times \exp -\frac{(x-x_2)}{L}$$

The total probability is given by integrating over all values of x_1 and x_2 from 0 to x with the restriction that $x_1 < x_2$; or alternatively, we can integrate them without restriction, and divide the result by two to account for the spurious doubling which arises in this method from the interchange of x_1 and x_2. In either case, we obtain:

$$\begin{aligned} Q_2(x) &= \left(\frac{1}{L^2}\right)\int_0^x dx_2 \int_0^x dx_1 \\ &\times \exp\left[\frac{-x_1}{L}-\frac{(x_2-x_1)}{L}-\frac{(x-x_2)}{L}\right] \\ &= \left(\frac{1}{L^2}\right)e^{-x/L}\int_0^x dx_2 \int_0^{x_2} dx_1 \\ &= \left(\frac{1}{L^2}\right)e^{-x/L}\int_0^x x_2\, dx_2 \\ &= \frac{1}{2}\left(\frac{x}{L}\right)^2 e^{-x/L}. \end{aligned} \tag{3.46}$$

Carrying on similarly, we can find the general result for the probability of precisely n collisions occurring in the distance x:

$$Q_n(x) = \frac{1}{n!}\left(\frac{x}{L}\right)^n e^{-x/L} \tag{3.47}$$

Notice that the total probability of all possible numbers of collisions in a given distance is unity, as it should be, independent of the distance:

$$\begin{aligned} Q_n(x) &= \sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{x}{L}\right)^n e^{-x/L} \\ &= e^{+x/L}\cdot e^{-x/L} = 1 \end{aligned} \tag{3.48}$$

Hence the $Q_n(x)$ constitute a discrete distribution for each x; we may use them, for example, to calculate the

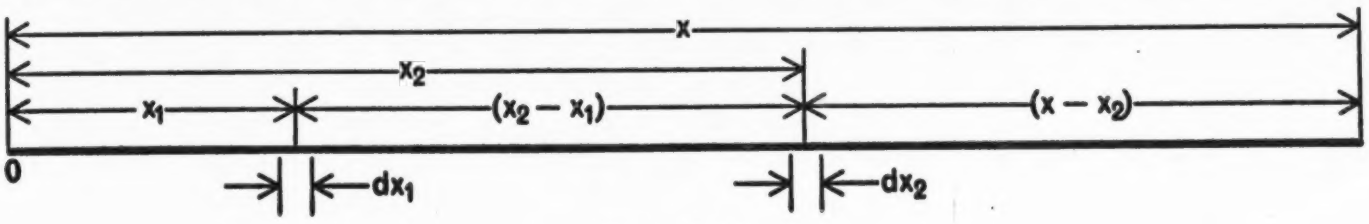


Fig. 3.23

mean number of collisions to be expected in a distance x:

$$\langle n \rangle = \sum_{n=0}^{\infty} n Q_n(x)$$

$$= \sum_{n=0}^{\infty} \frac{n}{n!} \left(\frac{x}{L}\right)^n e^{-x/L}$$

$$= \frac{x}{L} \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \left(\frac{x}{L}\right)^{n-1} e^{-x/L}$$

$$= \frac{x}{L} e^{+x/L} e^{-x/L} = \left(\frac{x}{L}\right). \quad (3.49)$$

Reasonably enough, the mean number of collisions in a given distance is found by dividing the distance by the mean free path. Figure 3.24 shows sketches of the Q_n plotted as functions of x; the predominance of small numbers of collisions for short paths, and the gradually increasing importance of larger numbers as the paths become longer, is evident from the curves.

It is important to recognize, however, that $Q_n(x)$ is <u>not</u>, as it stands, a continuous distribution with respect to x, of the type we have been studying heretofore in this chapter; it is neither a probability density, nor a cumulative distribution. We <u>could</u> ask: What is the probability that the n'th collision takes place between x and (x + dx)?, or what is the probability that the nth collision occurs between 0 and x? These functions <u>would</u> be continuous distributions in x; a density, and a cumulative distribution respectively.

Let us denote by $p_n(x)\,dx$ the probability that the nth collision takes place between x and (x + dx); it must be given by the product of the probability that precisely (n − 1) collisions take place in the distance x, by the probability that one collision takes place in dx. But the former is $Q_{n-1}(x)$ and the latter is (dx/L); hence

$$p_n(x)\,dx = Q_{n-1}(x) \left(\frac{dx}{L}\right)$$

$$= \frac{1}{(n-1)!} \left(\frac{x}{L}\right)^{n-1} e^{-x/L} \left(\frac{dx}{L}\right). \quad (3.50)$$

We see that the first one of the family, $p_1(x)$, is identical with the p(x) of the previous section Eq. (3.43), as it should be, and that all the p_n's satisfy the required normalization condition:

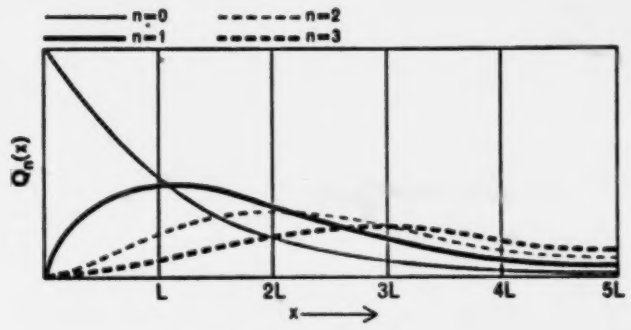


Fig. 3.24

$$\int_0^{\infty} p_n(x)\,dx$$

$$= \frac{1}{(n-1)!} \int_0^{\infty} \left(\frac{x}{L}\right)^{n-1} e^{-x/L} \left(\frac{dx}{L}\right)$$

$$= \frac{1}{(n-1)!} \cdot (n-1)! = 1 \quad (3.51)$$

Hence we may calculate averages in the usual ways; for example, the mean distance to the nth collision is given by:

$$\langle x \rangle = \int_0^{\infty} x\, p_n(x)\,dx$$

$$= \frac{L}{(n-1)!} \int_0^{\infty} \left(\frac{x}{L}\right)^n e^{-x/L} \left(\frac{dx}{L}\right)$$

$$= \frac{L\,n!}{(n-1)!} = nL. \quad (3.52)$$

This result is of course closely related to the result in Eq. (3.49), but they are by no means identical. Imagine regarding the two formulas as prescriptions for an experimental determination of the mean free path L; then the first says: "Take many sections of path, each of length x; find the average number of collisions in each, and divide it into x to obtain L." The second says: "Pick a number n of collisions; then measure many times the path required to give that number. Divide n into the average path length to obtain L."

Similarly, the two distributions $p_n(x)$ and $Q_n(x)$ are closely related, but their difference must be clearly understood. $Q_n(x)$ is a discrete distri-

bution with n as the index enumerating the various alternatives; x enters as a parameter, fixed in advance, though capable of taking on a continuous range of values from 0 to ∞. It answers the question "Given x, what is the distribution of n?" But $p_n(x)$ is a continuous distribution in x - a probability density function - in which n enters as a parameter, fixed in advance, which may have any integer value. It answers the question "Given n, what is the distribution of x?"

The discrete distribution $Q_n(x)$ is known to statisticians as the Poisson distribution. It is related to the asymmetric binomial distribution of section 2.4 in a way which can be understood by using a somewhat different approach to the collision problem. Imagine the length x of path divided into a large number M of segments; M is to be taken large enough so that (x/M), the length of each segment, is small compared to the mean free path L. Then for each segment two alternatives are possible: either there was a collision in that segment, or there was not; the former alternative has the probability (length of segment)/(mean free path), i.e., p = x/ML; the latter has probability q = 1 − p = 1 − (x/ML). Since there are M segments, the distribution of n, the number of occurrences of collisions, is given by the asymmetric binomial distribution (Equation (2.27):

$$p_n = \frac{M!}{n!\ (M-n)!}\ p^n\ q^{M-n} \qquad (3.53)$$

Now p is very small; furthermore, we are interested in the limit in which $M \to \infty$. With this in mind, we rearrange the terms of Eq. (3.53) as follows:

$$p_n = \frac{M(M-1)(M-2)\cdots(M-n+1)}{n!} \times \left(\frac{x}{ML}\right)^n \left(1-\frac{x}{ML}\right)^{M-n} = \frac{1}{n!}\left(\frac{x}{L}\right)^n \left[1-\left(\frac{x}{ML}\right)\right]^M \times \left\{\frac{(1-1/M)(1-2/M)\cdots[1-(n-1)/M]}{[1-(x/ML)]^n}\right\} \qquad (3.54)$$

Now for fixed n, the term in the curly brackets approaches unity as $M \to \infty$; and in the same limit, the factor $[1-(x/ML)]^M$ approaches $e^{-x/L}$, according to one of the definitions of the exponential function. We have therefore in the limit:

$$p = \frac{1}{n!}\left(\frac{x}{L}\right)^n e^{-x/L}, \qquad (3.55)$$

and we see that this is identical with the $Q_n(x)$ deduced in a different way previously. Thus the Poisson distribution is a limiting case of the binomial distribution in which one of the alternatives has an extremely small probability. As such it has many applications other than the one we have been discussing here.

3.5 GAUSSIAN DISTRIBUTION; ERRORS

There exists a different approximation to the binomial distribution from the one considered in the previous section, which is of great importance. It is the limit in which M is large, and the probabilities in the neighborhood of the maximum of the distribution, which occurs at M/2 in the symmetric case, are of primary interest. Furthermore, we are often interested in regarding the index k of the distribution as continuous rather than as discrete; this can be made meaningful when M is sufficiently large, when changes of several units in k correspond to only very slight changes in the associated probability p_k. Let us see how to handle this approximation. We will use the symmetric case for simplicity; the general asymmetric case can be done similarly.

We begin by considering the ratio of the probability for arbitrary index k to the probability at the maximum, which occurs at k = M/2. According to Eq. (2.20), this is given by:

$$\frac{p_k}{p_{M/2}} = \frac{[(M/2)!]^2}{k!\ (M-k)!}. \qquad (3.56)$$

It will be useful to introduce an in-

dex m which is measured from the peak of the curve:

$$m = k - (M/2) \qquad (3.57)$$

Then we have

$$\frac{p_m}{p_{M/2}} = \frac{(M/2)!\ (M/2)!}{[(M/2)+m]!\ [(M/2)-m]!}$$

$$= \frac{(\frac{1}{2}M)(\frac{1}{2}M-1)(\frac{1}{2}M-2)\ldots(\frac{1}{2}M-m+1)}{(\frac{1}{2}M+m)(\frac{1}{2}M+m-1)(\frac{1}{2}M+m-2)\ldots(\frac{1}{2}M+1)}$$

$$= \frac{1[1-(2/M)][1-(4/M)]\ldots[1-2(m-1)/M]}{[1+(2/M)][1+(4/M)]\ldots[1+(2m/M)]} \qquad (3.58)$$

We have written it this way to show that if m is much less than M - which is the case in the central part of the distribution that we want to approximate to - then each factor in both the numerator and the denominator is close to unity. However, there are many factors, for even if m is much less than M, it can be itself large compared with unity. It will be helpful, since we have a product of many factors, to take the logarithm, since the logarithm of the product of several factors is equal to the sum of their logarithms. Doing this, we obtain

$$\ln\left(\frac{p_m}{p_{M/2}}\right) = \sum_{j=1}^{m-1} \ln\left[1-\left(\frac{2j}{M}\right)\right] - \sum_{j=1}^{m} \ln\left[1+\left(\frac{2j}{M}\right)\right]$$

$$= \sum_{j=1}^{m-1}\left[\frac{2j}{M} - \frac{1}{2}\left(\frac{2j}{M}\right)^2 \ldots\right] - \sum_{j=1}^{m}\left[\frac{2j}{M} - \frac{1}{2}\left(\frac{2j}{M}\right)^2 \ldots\right]$$

Here we have used the Taylor expansion of the function ln (1 + x) for small x. Now using the summation formulae

$$\sum_{j=1}^{m-1} j = m(m-1)/2$$

and

$$\sum_{j=1}^{m} j = m(m+1)/2,$$

and neglecting all but the first term in each sum because the higher terms give rise to expressions proportional to m^3/M^2, which is much less than (m^2/M), we obtain.

$$\ln \frac{p_m}{p_{M/2}} = -\frac{(2/M)m(m-1)}{2} - \frac{(2/M)m(m+1)}{2}$$

$$= -2m^2/M,$$

or:

$$p_m/p_{M/2} = \exp(-2m^2/M)$$

and

$$p_m = p_{M/2} \exp(-2m^2/M). \qquad (3.59)$$

If we reexpress this in terms of the original index k, we have:

$$p_k = p_{M/2} \exp\{-2[k-(M/2)]^2/M\}. \qquad (3.60)$$

This is evidently of the Gaussian form defined in Eq. (3.12), except that it is a discrete distribution in an index k taking on integer values rather than a probability density in a continuous variable. We may take the transition to that form by observing that if M is large enough so that there exist intervals Δk which, although they contain several values of k (that is, they include several consecutive integers), they are still sufficiently small compared to M so that p_k changes very little in that interval, then the probability that the index lies between k and $(k + \Delta k)$ is simply $p_k \Delta k$. If we now rewrite this in the style of a density function, we have:

$$p(k)\,dk = p_{max} \exp\{-2[k-(M/2)]^2/M\}\,dk, \qquad (3.61)$$

which is now precisely of the form of a Gaussian density function, with mean M/2 and standard deviation $\sigma^2 = M/4$.

Notice that these values are identical with those found in Chapter 2, Eqs. (2.22) and (2.24). The constant p_{max} which occurs in the distribution has of course a known value; but it is simpler to observe that the normalization condition requires it to be:

$$p_{max} = \sqrt{2/\pi M}. \qquad (3.62)$$

Thus the Gaussian distribution appears as a limiting form of the binomial distribution, appropriate for large M and in the neighborhood of the maximum of the distribution.

One of the areas of application of the Gaussian (or "normal," as the statisticians call it!) distribution is in the discussion of experimental error. It is often said that the measured values of any physical quantity, if a sufficient number are accumulated, form a Gaussian or normal distribution about the "true" value. Leaving aside any deeper discussion of the question of what one means by the "true" value, let us simply suppose that there does, in fact, exist a correct value q_0 (possibly determined by using more refined apparatus) of the quantity being measured, and let us inquire what might be the cumulative effect of various possible sources of error in the measurement. A very crude model of the process is the following: Suppose there are M sources of error in the experiment, and each gives rise to an error $+\epsilon$ or $-\epsilon$ with equal probability. The actual result of an individual measurement of the quantity q will then be:

$$q = q_0 + (k - j)\epsilon, \qquad (3.63)$$

where k is the number of times $+\epsilon$ occurs, and j is the number of times that $-\epsilon$ occurs. Now $k + j = M$; hence

$$q = q_0 + (2k - M)\epsilon. \qquad (3.64)$$

Now the repetition M times of the choice $+\epsilon$ or $-\epsilon$ gives rise to a binomial distribution for k, the number of $+\epsilon$ choices; hence

$$p_k = \frac{M!}{k!\,(M-k)!}\,\frac{1}{2^M} \qquad (3.65)$$

is the probability that an individual measurement will yield a value of q corresponding (through Eq. (3.64) to a given k. Rewriting this in the approximation derived earlier, we have

$$p(k)\,dk = \sqrt{2/\pi M}\,\exp\left\{-2[k - (M/2)]^2/M\right\}\,dk.$$

Now let us use the results of section 3.2 to convert from a distribution in k to a distribution in q:

$$p(k) = \sqrt{2/\pi M} \times \exp - \left[(2/M)(q - q_0)/2\epsilon\right]^2\,dq/2\epsilon.$$

Hence the probability density function for q is given by

$$p(q) = \sqrt{1/2\pi M\epsilon^2}\,\exp\left[-(q - q_0)^2/2M\epsilon^2\right], \qquad (3.66)$$

so that the distribution of q values is is indeed Gaussian, with mean q_0 and standard deviation $\sqrt{M}\,\epsilon$.

This is, of course, a very crude and unrealistic model of the "error" problem; nevertheless, it does happen that even under considerably less restrictive assumptions, one arrives at a Gaussian distribution of errors. This does not, of course, guarantee that in any particular experiment the distribution of values obtained necessarily follows the "normal" distribution; but it is indeed observed to hold in sufficiently many situations to make it a very useful first assumption.

3.6 JOINT DISTRIBUTIONS; MAXWELLIAN VELOCITY DISTRIBUTION

In the discrete case, we studied the joint distribution arising in the problem of assignment of molecules to several cells into which a volume of gas had been divided. The collection of integers $k_1, k_2, k_3, \ldots k_N$ giving

the number of molecules occupying each cell was found to occur with a probability $p(k_1, k_2, \ldots k_N)$ which depended jointly on all the k 's; this was called therefore a "joint distribution." In the continuous case the analogous situation occurs; we often have to deal with joint distributions of several variables. An example to which we will return later is the distribution of the velocity components of a molecule in a gas. Each molecule has a velocity vector specified by the three components along the three axes of reference, say v_x, v_y, and v_z. We need to know the probability that simultaneously the x component has a value lying between v_x and $(v_x + dv_x)$, the y component has a value between v_y and $(v_y + dv_y)$, and the z component has a value between v_z and $(v_z + dv_z)$. Or we may be interested in its speed and direction, and therefore inquire about the probability of simultaneously finding its speed between v and (v + dv) while its direction lies within the solid angle defined by the intervals from θ to $(\theta + d\theta)$ and ϕ to $(\phi + d\phi)$ of the spherical polar angles defining its direction with respect to a fixed polar axis.

In general, we define the joint probability density of two quantities by the function of two variables $p(x,y)$ such that $p(x,y)\,dx\,dy$ gives the probability that the first lies between x and (x + dx) and the second lies between y and (y + dy). The generalization to more variables is made in the obvious way: $p(x,y,z)\,dx\,dy\,dz$ for three, and so on. The density must satisfy a normalization condition:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x,y)\,dx\,dy = 1, \qquad (3.67)$$

which assures us that the two quantities are certain to be found somewhere in their range. The probability that x and y lie within any region R of the x-y plane is given by the integral:

$$\iint_R p(x,y)\,dx\,dy,$$

where the integration symbol $\iint_R$ means a two-dimensional integral over the region R.

From such a joint distribution, the distribution of either variable separately may be found; for the probability that the first one lies between x and (x + dx), irrespective of the value of y, is given by

$$dx \int_{-\infty}^{\infty} p(x,y)\,dy. \qquad (3.68)$$

This follows by taking R to be the whole region corresponding to x lying in the interval dx while y is anywhere; this is simply a strip of width dx running parallel to the y axis, and the corresponding integration is just what is expressed in Eq. (3.68). Similarly the probability that the second variable lies between y and (y + dy), irrespective of the value of x, is given by

$$dy \int_{-\infty}^{\infty} p(x,y)\,dx. \qquad (3.69)$$

Each of these is properly normalized by virtue of Eq. (3.67).

Mean values of functions of x and y are defined in a way which is the appropriate generalization of the single-variable Eq. (3.8) to the case of two variables:

$$\langle f(x,y)\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,p(x,y)\,dx\,dy. \qquad (3.70)$$

In addition to the means $\langle x\rangle, \langle y\rangle$, $\sigma_x^2 = \langle x^2\rangle - \langle x\rangle^2$, and $\sigma_y^2 = \langle y^2\rangle - \langle y\rangle^2$, which we have used frequently, a new average called the "correlation coefficient" and defined by the equation:

$$r = (\langle xy\rangle - \langle x\rangle\langle y\rangle)/\sigma_x\sigma_y \qquad (3.71)$$

becomes important; it is an indicator of the degree to which the two variables are independent of one another. They are independent if $p(x,y)$ is of the form of a product $p_1(x)\,p_2(y)$; for then the probability of the joint event "x in dx and y in dy" is simply the product of the probabilities of the events "x in dx" irrespective of y, and "y in dy," irrespective of x. Since the probability of joint events is the

product of the individual probabilities only if the events are independent, the product form $p_1(x)\ p_2(y)$ for the joint probability indicates independence of x and y. We see that r will vanish in this case, for then we will have $\langle xy \rangle = \langle x \rangle \langle y \rangle$. If, however, x and y are "correlated," then r need not vanish. If, for example, there is a tendency for positive x to occur with positive y, and negative x with negative y, then r will be positive.

Let us give some simple examples of joint distributions.

3.6.1 Uniform Distribution Over a Rectangular Area

Suppose a point is to be picked "at random" from the rectangle in the x-y plane defined by $0 \le x \le L_1$, $0 \le y \le L_2$; what joint probability density function describes this? "At random" means here that the point is equally likely to be found in any two areas of the same size within the rectangle; hence p(x,y) is constant over the rectangle, and zero elsewhere. Bearing in mind the normalization condition, Eq. (3.67), we may write:

$$p(x,y)\,dx\,dy = \begin{cases} (dx\,dy)/L_1L_2 & \begin{cases} 0 \le x \le L_1 \\ 0 \le y \le L_2 \end{cases} \\ 0 & \text{elsewhere} \end{cases} \tag{3.72}$$

In this case the variables are independent, since p(x,y) can be written as the product $p_1(x)\ p_2(y)$, where p_1 and p_2 are each of the form of the uniform one-dimensional distribution given by Eq. (3.9). The various averages can be found easily.

3.6.2 Uniform Angular Distribution

Directions in space may be specified by the spherical polar coordinates θ and ϕ which give the position of the point on a sphere at which a ray from the origin intersects a sphere. By a "uniform angular distribution" we mean that the intersection point is equally likely to be found in equal areas on the sphere; therefore to express the probability density function in terms of θ and ϕ we need to find what fraction of the total area of the sphere corresponds to the area on the sphere defined by the range $d\theta$ of θ and $d\phi$ of ϕ. Figure (3.25) shows the relevant area; we see that it has dimensions $R\,d\theta$ by $R \sin\theta\,d\phi$. The area of the sphere is $4\pi R^2$; hence the ratio of the two, which gives the probability that the point lies in the range of θ and ϕ specified, is given by:

$$p(\theta,\phi)\,d\theta\,d\phi = (R\,d\theta)(R \sin\theta\,d\phi)/4\pi R^2$$

$$= \sin\theta\,d\theta\,d\phi/4\pi. \tag{3.73}$$

Notice that here also the variables are independent. Furthermore, we may work out the distribution of θ alone; according to Eq. (3.68) this is given by:

$$d\theta \int_0^{2\pi} p(\theta,\phi)\,d\phi = (\sin\theta\,d\theta)/4\pi \int_0^{2\pi} d\phi$$

$$= 2\pi\,(\sin\theta\,d\theta)/4\pi$$

$$= \sin\theta\,d\theta/2,$$

which agrees with the result of section 3.1.2, Eq. (3.11).

Just as for distributions in one variable, the question of change of variables arises for joint distributions. We saw in section 3.2 how to handle such a change; the method is simply to substitute the change of variable, both in the density function and in the differential which accompanies it. The coefficient of the differential of the new variable is then the density function in the new variable.

For joint distributions in two or more variables an analogous procedure is appropriate, but one must remember the method of transforming an element of area or volume from one set of variables to another. If p(x,y) dx dy is the density function in x and y, and if they are each functions of two new

variables u and v whose distribution is required,

$$x = f(u,v)$$

$$y = g(u,v),$$

then in the calculus of two or more variables we learn that the method of transforming elements of area from the x-y plane to the u-v plane is this:

$$dx\ dy = \begin{vmatrix} \partial f/\partial u & \partial f/\partial v \\ \partial g/\partial u & \partial g/\partial v \end{vmatrix} du\ dv. \quad (3.74)$$

The determinant of the partial derivatives which enters is called the Jacobian determinant of the transformation, and is often denoted by J(f,g/u,v). Using this notation, we may write the transformation of the density function thus:

$$p(x,y)\ dx\ dy = p[f(u,v),g(u,v)] \times J(f,g/u,v)\ du\ dv, \quad (3.75)$$

so that the new density function q(u,v) is given by:

$$q(u,v) = p[f(u,v),g(u,v)]\ J(f,g/u,v). \quad (3.76)$$

If, for example, we wish to transform from (x,y) to polar coordinates in a plane (r,θ), for which $x = r\cos\theta$, $y = r\sin\theta$, we find that $dx\ dy = r\ dr\ d\theta$, and $p(x,y)\ dx\ dy = p(r\cos\theta, r\sin\theta)\ r\ dr\ d\theta$, so that

$$q(r,\theta) = r\ p(r\cos\theta, r\sin\theta).$$

In a similar fashion the transformation from a distribution in three variables (x,y,z) to one in the corresponding spherical polar coordinates (r,θ,ϕ), for which the equations relating the coordinates are

$$x = r\sin\theta\cos\phi,$$

$$y = r\sin\theta\sin\phi,$$

$$z = r\cos\theta,$$

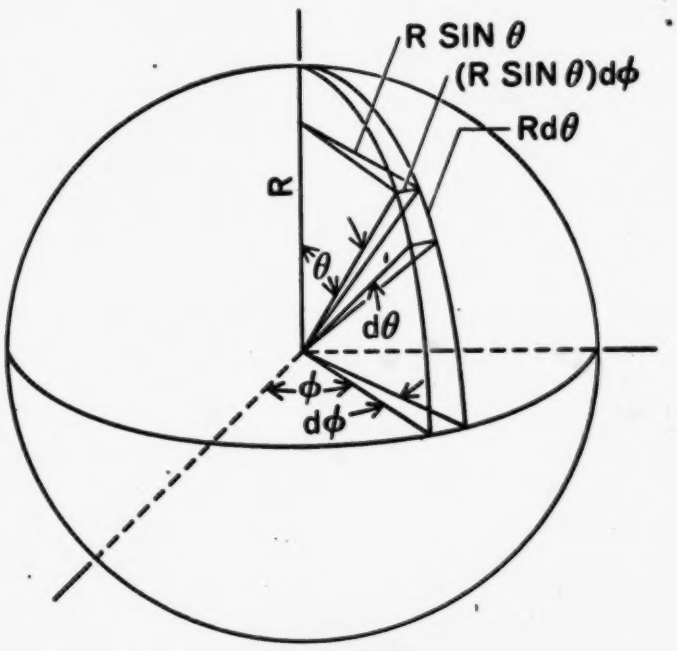


Fig. 3.25

will result in a volume element transformation

$$dx\ dy\ dz = r^2 \sin\theta\ dr\ d\theta\ d\phi,$$

so that if the original density function is p(x,y,z), the transformed one is:

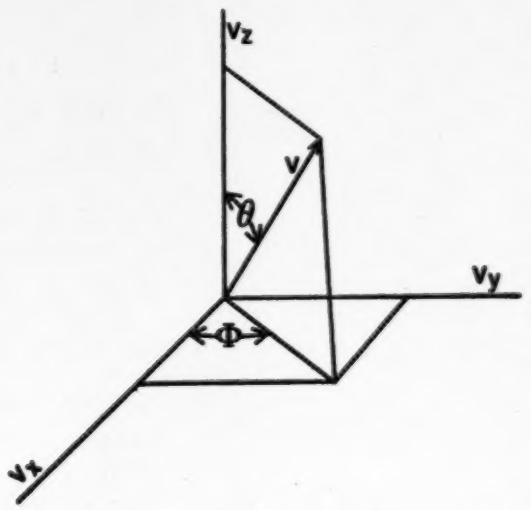
$$q(r,\theta,\phi) = p(r\sin\theta\cos\phi,\ r\sin\theta\sin\phi,\ r\cos\theta)r^2\sin\theta.$$

Notice that if, for example, the density function in the polar coordinates were required to be uniform in angle, we would have to require that p be of such a form that no angular dependence arise from it, since the $\sin\theta$ factor from the volume element already gives the form of angular dependence required by Eq. (3.73) to describe a uniform angular distribution. What this means is that p can depend on (x,y,z) only through the combination $(x^2 + y^2 + z^2)$, which is of course just r^2 and is independent of angle.

As a final example of a joint distribution, we will discuss briefly the distribution of molecular velocities in a gas. This distribution, which is very important in the kinetic theory of gases, is generally known as the "Maxwellian" distribution in honor of

James Clark Maxwell, who was one of the first physicists to explore the kinetic theory mathematically. A number of different derivations exist; we shall give here one similar to one of Maxwell's early discussions. It is not to be regarded as a rigorous proof, because a number of unproved assumptions will be made in the course of the argument. Nevertheless it is interesting, and it illustrates the use of arguments of a type much used in modern physics - arguments of symmetry.

Let us then denote by $p(v_x, v_y, v_z)$ the joint probability density for finding the x component of velocity of a molecule between v_x and $(v_x + dv_x)$, the y component between v_y and $(v_y + dv_y)$, and the z component between v_z and $(v_z + dv_z)$. It will be useful to imagine a "velocity space," that is, a set of Cartesian coordinates corresponding to each velocity component; a given point in this space corresponds to the velocity vector extending from the origin to the point. The point may also be described by spherical polar coordinates; (v,θ,ϕ) instead of (v_x, v_y, v_z); in this case v represents the speed, or magnitude of the velocity vector $(v^2 = v_x^2 + v_y^2 + v_z^2)$, and θ and ϕ are the polar angles corresponding to the direction of the velocity. Figure 3.26 shows the relation between the various coordinates.



$v_x = v \sin\theta \cos\phi$
$v_y = v \sin\theta \sin\phi$
$v_z = v \cos\theta$

Fig. 3.26

What general arguments can we give about the form of the distribution function? First: $p(v_x, v_y, v_z)$ should be of such a form that all directions of motion are equally probable for the molecule; we do not expect nature to have any preference for one direction over another. This means that the angular distribution should be uniform, in the sense in which we used the term earlier. This means that p must depend on the velocity components only through the speed v, since as we saw before, the transformed volume element in polar coordinates already describes a uniform angular distribution. We can express this result by writing

$$p(v_x, v_y, v_z) = G(v^2)$$

$$= G(v_x^2 + v_y^2 + v_z^2), \qquad (3.77)$$

where by G we mean a function of a single variable. Second: We will suppose that v_x, v_y and v_z are independent, that is, that the distribution of any one of them is the same for fixed values of the other two, regardless of the particular values chosen for the other two. As we saw earlier in this section, independence implies that the joint distribution is a product of individual distributions:

$$p(v_x, v_y, v_z) = F_1(v_x)\, F_2(v_y)\, F_3(v_z).$$

This is the weakest assumption of this method of arriving at the Maxwellian distribution. It is by no means obvious that if we examine those molecules with $v_y = v_z = 0$, we will find precisely the same distribution of values of v that we will find if we examine those molecules with, say, large values of v_y and and v_z. Nevertheless, the randomizing effect of the molecular collisions does indeed have the net effect of making the distribution of the components independent. It is not easy to demonstrate this in a simple way, so we shall simply assume it. Third: the individual distributions F_1, F_2, and F_3

must actually be identical, for the labeling of the axes is quite arbitrary; it clearly cannot matter which direction we happen to call x. The distribution of the x component must be the same as that of the y component. Thus:

$$p(v_x, v_y, v_z) = G(v_x^2 + v_y^2 + v_z^2) = F(v_x)\,F(v_y)\,F(v_z). \qquad (3.78)$$

The remarkable fact is that these assumptions completely determine the form of the distribution function. Notice first that if we set $v_y = v_z = 0$ in Eq. (3.78), we get a relation between the functions F and G:

$$G(v_x^2) = F(v_x)\,[F(0)]^2, \qquad (3.79)$$

which, when substituted in Eq. (3.78) for F, gives us a relation for the single function G:

$$G(v_x^2 + v_y^2 + v_z^2) = G(v_x^2)\,G(v_y^2)\,G(v_z^2)/[F(0)]^4. \qquad (3.80)$$

It will simplify matters a bit if we temporarily use (u,v,w) in place of the squares of the velocity components. If we also notice that from Eq. (3.79) we have $G(0) = [F(0)]^2$, and if we let $g(u) = G(u)/G(0)$, then Eq. (3.80) can be rewritten as:

$$g(u + v + w) = g(u)\,g(v)\,g(w). \qquad (3.81)$$

The only possible function satisfying this relation is an exponential function, as we see if we differentiate both sides with respect to u, and then set $u = v = 0$; this yields:

$$g'(w) = g'(0)\,g(w). \qquad (3.82)$$

The solution of this differential equation for which $g(0) = 1$ is:

$$g(w) = e^{-\alpha w}, \quad \alpha = -g'(0). \qquad (3.83)$$

Returning to the original notation and inserting the result into Eq. (3.76) we obtain the distribution function:

$$p(v_x, v_y, v_z) = G(v^2) = G(0)\,g(v^2)$$
$$= G(0)\exp -\alpha v^2$$
$$= G(0)\exp -\alpha(v_x^2 + v_y^2 + v_z^2) \qquad (3.84)$$

The constant G(0) is determined by the normalization condition:

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p\,dv_x\,dv_y\,dv_z$$
$$= G(0)\int_{-\infty}^{\infty}\exp(-\alpha v_x^2)\,dv_x$$
$$\times \int_{-\infty}^{\infty}\exp(-\alpha v_y^2)\,dv_y$$
$$\times \int_{-\infty}^{\infty}\exp(-\alpha v_z^2)\,dv_z$$
$$= G(0)\,(\pi/\alpha)^{3/2}.$$

Hence, the final form of the distribution function is:

$$p(v_x, v_y, v_z) = (\alpha/\pi)^{3/2}$$
$$\times \exp -\alpha(v_x^2 + v_y^2 + v_z^2). \qquad (3.85)$$

This is the Maxwellian distribution; we see that it is "Gaussian" in each velocity component. The constant α remains undetermined; evidently it has to do with the width of the distribution, that is, the mean square velocity. In the kinetic theory one shows that it must be inversely proportional to the absolute temperature of the gas.

We may write the distribution in the polar form; using the fact that the volume element becomes $v^2 \sin\theta\,dv\,d\theta\,d\varphi$ in spherical polar coordinates, the distribution becomes:

$$q(v, \theta, \varphi) = (\alpha/\pi)^{3/2}\exp(-\alpha v^2)$$
$$\times v^2 \sin\theta. \qquad (3.86)$$

By integrating over the polar angles,

we can get the distribution of speed alone:

$$q(v) = \int_0^{2\pi}\int_0^{\pi} q(v, \theta, \phi)\, dv\, d\theta\, d\phi$$

$$= (a/\pi)^{3/2} \exp(-av^2)v^2 \times \int_0^{2\pi}\int_0^{\pi} \sin\theta\, d\theta\, d\phi$$

$$= 4a^{3/2}/\pi^{1/2}\, v^2 \exp -av^2. \qquad (3.87)$$

From these results we can work out the various averages of interest: mean speed, mean square speed, mean square of a velocity component, and so on. The role of the Maxwellian distribution of velocities is fundamental in many fields of physics, and the generalization to the Maxwell-Boltzmann distribution which is made in statistical mechanics underlies the whole of thermal physics.

PROBLEMS

3.1 For the case of scattering of a particle with charge q_1 at a fixed scattering center with charge q_2, by virtue of the electrostatic interaction between them, the relation between impact parameter b and scattering angle θ is given by classical dynamics to be $b = (q_1 q_2/mv^2)\cot(\theta/2)$, where m is the mass and v the speed of the scattered particle. Find the angular distribution of the scattered particles. (This is the "Rutherford scattering" from which Rutherford deduced the existence of small, massive, charged nuclei in atoms.)

3.2 Looking out the window, I see objects falling past me. Measuring their speeds, I find that they are distributed uniformly between 40 ft/sec and 80 ft/sec.

(a) Assume they are being dropped from rest at various heights above me. What is the distribution of heights required to give my observed distribution of speeds?

(b) Assume instead that they are all being dropped from 25 feet above me, but with various initial speeds. What distribution of initial speeds will account for my observed distribution of speeds?

3.3 For each of the distributions discussed in section 3.4, find the standard deviation, for $Q_n(x)$ this will mean finding $\langle n^2\rangle$, and for $P_n(x)\,dx$, finding $\langle x^2\rangle$. In each case compare the standard deviation with the mean value itself, and use this result to formulate instructions for the experimenter interested in measuring the mean free path to an accuracy of about 5% in the way discussed immediately following Eq. (3.52).

3.4 The counts produced by a Geiger counter exposed to a steady source which gives an average rate of one count in time T, form a time sequence governed by the same Poisson distribution that governs the molecular paths discussed in the text; it is only necessary to substitute t for x and T for L. What is the probability that in a time 3T there occur (a) 2 or less counts? (b) Exactly 3 counts? (c) 4 or more counts?

3.5 The probability that the first collision takes place between 0 and x is given by $P(x)$, Eq. (3.42). Call this $P_1(x)$; then find $P_2(x)$, the probability that the second collision takes place between 0 and x. (Hint: This is the cumulative distribution associated with the density $p_2(x)$.) At what distance will one have 95% probability

that the first collision has occurred? The second?

3.6 Suppose a needle of unit length is dropped onto a plane marked with parallel lines unit distance apart. It falls "at random" on the plane. Take this to mean that the joint probability of finding the center of the needle at a distance between x and (x + dx) to the right of a line and of finding the angle which the needle makes with lines to lie between θ and (θ + dθ), is uniform over the range $0 \le x \le 1$ and $-\pi/2 < \theta < +\pi/2$. Find the probability that the needle intersects a line, and the probability that it does not intersect a line. (Hint: You will need to find the area R in the x-θ plane corresponding to the required conditions.)

3.7 Suppose x is uniformly distributed between 0 and 1, and y is also uniformly distributed between 0 and 1, and they are independent of one another. What is the joint probability density for x and y? Find the region R of the x-y plane for which the average (x+y)/2 lies between 0 and s. What is the probability P(s) that the average lies in this region? What is the probability density function p(s) ds of the average? Find σ^2 for the original distribution of x (or y), and find σ^2 for the distribution of the average s. Speculate on the probable behavior of σ^2 as the number of identically distributed independent variables averaged increases.

3.8 (a) Show that the mean square of each component of the velocity of the molecules in a gas is one third of the mean square speed. (Hint: It is not necessary to perform any integrals; use the relation between v^2 and the squared components, the definition of the mean, and a symmetry argument.)

(b) From statistical mechanics we know that the mean kinetic energy $\frac{1}{2}m\langle v^2\rangle$ of the molecule in a gas must be equal to 3kT/2, where k is Boltzmann's constant and T is the absolute temperature. Use this to establish the relation between the constant a appearing in the Maxwellian distribution and the absolute temperature T.